MCPRO+ 2.6

User Manual

Schrödinger Press

June 2009

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

Links to other locations in the current document or to other PDF documents are colored like this: Document Conventions.

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

File name, path, and environment variable syntax is generally given with the UNIX conventions. To obtain the Windows conventions, replace the forward slash / with the backslash \ in path or directory names, and replace the $ at the beginning of an environment variable with a % at each end. For example, `$SCHRODINGER/maestro` becomes `%SCHRODINGER%\maestro`.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Introduction

1.1 About MCPRO+

MCPRO is a well-validated simulation package from Prof. Bill Jorgensen that performs Monte Carlo statistical mechanics simulations of ligands, peptides, proteins, and nucleic acids in the gas phase or in solution. Accurate free energy changes are computed from statistical perturbation theory to generate relative or absolute free energies of binding for protein/ligand complexes, reaction profiles, potentials of mean force, and free energy profiles for rotation about dihedral angles.

MCPRO+ is a simulation package built around the MCPRO program. It includes a number of enhancements and a powerful user interface to MCPRO for protein-ligand simulations. Along with the use of the Schrödinger job control facility, the MCPRO+ interface makes setting up, running and monitoring of MCPRO simulations simple. Key features include:

- Interface for non-expert users to reliably calculate relative binding free energies for protein/ligand complexes via highly-accurate free energy perturbation (FEP) theory.
- Prediction of protein–ligand binding free energies using linear response methodology.
- Energy minimization in Cartesian or internal coordinates.
- Full range of aqueous solvation options including periodic explicit solvent boxes or solvent caps and the GB/SA implicit solvent model.
- Automatic atom typing of macromolecular species with OPLS_2001 or OPLS_2005 force fields.
- Boltzmann-weighted trajectories from NPT or NVT Monte Carlo sampling, which can be used analogously to molecular dynamics trajectories.
- Integration with the Schrödinger Job Control facility, which provides job distribution across multiple CPUs and built-in job monitoring and job restarting.

MCPRO+ has been designed to be largely backward compatible with standard MCPRO input Z-matrix and parameter files and can be run using the shell scripts that have historically governed MCPRO job execution with only minor changes.

MCPRO+, like MCPRO, requires a Z-matrix as input for the system being simulated. Manual generation or editing of such a Z-matrix is error prone and time-consuming. MCPRO+ intro-

duces a number of features that greatly simplify job setup, the most important of which is the concept of a model system and a graphical interface for its creation. A model system consists of the Z-matrix of the system to be simulated, including atom parameter information, and a Maestro formatted structure file of the same structure. With the model system, you can view input and output structures from MCPRO+ simulations in Maestro using the Maestro files, and use them in other applications. Model systems also enable easy reuse of previously generated Z-matrices, such as when forming a set of protein-ligand complexes, and simplifies treatment of protein systems with cofactors and metal ions. Use of a model system allows you to merge additional structures into a Z-matrix while keeping the same frame of reference. Model system generation from Maestro is described in Chapter 3.

Another feature of MCPRO+ that facilitates its use is the MCPRO+ input file. This file contains the information that is contained in the MCPRO parameter files and shell scripts, in a simple keyword=value pair format. The file is divided into blocks in which options set for the entire simulation and per sampling block parameters can be adjusted. Using a reasonable set of defaults for most MCPRO+ options makes it unnecessary to include the settings for all parameters in the input file. Instead only those that differ from the defaults or are otherwise important to the user are included. To execute a Monte Carlo sampling run with MCPRO+, only three files are necessary: the two files that constitute the model system and the input file. The input file is then converted into a traditional shell script that governs the execution of the MCPRO executable. The format of the input file is described in Chapter 11.

1.2 The MCPRO+ Panels

The MCPRO+ panels provide an interface for setting up and running MCPRO+ jobs. The controls in the panel are configured differently for each kind of MCPRO+ task. MCPRO+ supports the following tasks:

- Minimization
- Free-energy perturbation
- Prediction of relative binding affinities
- Monte Carlo simulations
- Linear response simulations
- Examination of structure-activity relationships

Each of these tasks is described in a separate chapter of this manual.

To open the MCPRO+ panel for a particular task, choose the task from the MCPRO+ submenu of the Applications menu.

The main MCPRO+ panels contain a set of tabs and some action buttons. Each panel contains the Create Model Systems tab and the Solvation tab, and the third tab is task-dependent. For

information on the controls in a tab, see the help topic for the tab, which you can open by clicking Help when the tab is displayed. The use of these controls is also described in later chapters in this manual.

The Create Model Systems tab is wizard-like, and takes you through a series of steps to create the model system required for the calculation. Once the model system is defined, the other tabs become available, and you can set the rest of the parameters for the calculation. Creating a model system is described in Chapter 3.

The action buttons consist of a Start button, which opens a Start dialog box for setting up and submitting the job, a Write button for writing out input files, and a Reset button for clearing the settings and reverting them to the defaults.

Two other panels provide interfaces for the mutation of ligand side chains and ring atoms in a protein-ligand complex by FEP. These panels perform the model system creation automatically, so that you only need to provide the structures, select the mutation sites and the desired functional groups.

1.3 Running Schrödinger Software

To run any Schrödinger program on a UNIX platform, or start a Schrödinger job on a remote host from a UNIX platform, you must first set the `SCHRODINGER` environment variable to the installation directory for your Schrödinger software. To set this variable, enter the following command at a shell prompt:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

Once you have set the `SCHRODINGER` environment variable, you can start Maestro with the following command:

```
$SCHRODINGER/maestro &
```

It is usually a good idea to change to the desired working directory before starting Maestro. This directory then becomes Maestro's working directory. For more information on starting Maestro, including starting Maestro on a Windows platform, see Section 2.1 of the *Maestro User Manual*.

To run remote jobs, you must have access to a hosts file, named `schrodinger.hosts`, that lists hosts on which Schrödinger software is installed for execution. Details of setting up this file can be found in Chapter 6 of the *Installation Guide*.

MCPRO+ simulations usually run many subjobs. If you want to terminate the simulation, you must use the Job Control facility to kill the job and its subjobs. Do not use UNIX commands for this purpose, as they will not perform the necessary cleanup. For more information on using the Job Control facility for job management, see Chapter 3 of the *Job Control Guide*.

1.4 Citing MCPRO+ in Publications

The use of this product should be acknowledged in publications as:

MCPRO+, version 2.6, Schrödinger, LLC, New York, NY, 2009.

Theoretical Background

2.1 Statistical Mechanics Simulations

Simulations of many-particle systems commonly use Monte Carlo (MC) statistical mechanics or molecular dynamics (MD). Both methods typically employ a classical potential energy function for bond stretching, angle bending, torsions, and non-bonded interactions.

A MC simulation generates a new configuration by a set of random motions. The difference in energy between the new and the old configuration is used as a selection criterion by the Metropolis algorithm, which enforces a correct Boltzmann distribution of energies for the system at the desired temperature, and the procedure is then iteratively repeated. The resultant ensemble of configurations can be used to compute the structure and thermodynamic properties of the system at the specified conditions. Differences in free energies between two similar states of the system can also be calculated through statistical perturbation theory. In a MD simulation, on the other hand, the forces from the energy components in the three Cartesian directions acting on each atom are evaluated. The accelerations are calculated from these forces, and a trajectory is generated by the repeated numerical integration of the equations of motion over a period of time. The trajectory constitutes an ensemble of configurations that is formally equivalent to an ensemble generated via a MC simulation and is used in the same fashion.

This difference in methodology has significant implications in the use and application of the methods. The major algorithmic difference is that MC sampling can readily use internal coordinates, while MD works with independent atomic motions in Cartesian space. The consequences of this difference are many; in MD sampling is easier for large, flexible molecules since any coupling between different degrees of freedom is automatically handled by the algorithm. On the other hand, the integration time step has to be very small in order to avoid numerical instability. There is a certain inefficiency in the sampling since motions that go uphill in energy are usually reversed by opposing restoring forces. Also, the algorithm makes the selective freezing of unimportant degrees of freedom cumbersome. In a MC calculation, the freezing of selected internal coordinates is easily accomplished by simply not allowing variations in the specified bonds, angles, or dihedrals; however, sampling for large flexible molecules can be problematic since degrees of freedom are treated, for the most part, as being completely uncoupled. An additional advantage of MC is that since the motion is by nature random it is possible to sample large regions of configurational space, by stepping over potential energy barriers. In a paper comparing the efficiencies of MD and MC in simulations of

liquid hexane (W. J. Jorgensen; J. Tirado-Rives, J. Phys. Chem. **1996**, *100*, 14508), MC was found to be 2–4 times more efficient for conformational sampling.

Summaries on the theory and computations can be found in the following paper along with references to more detailed presentations: W. L. Jorgensen, J. Phys. Chem. **1983**, *87*, 5304. See also M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids; Oxford University Press: London, 1987.

2.2 Energy and Free Energy Evaluation

In MCPRO, the total potential energy of a system is given by

$$E = E_{SS} + E_{SX} + E_{XX} + E_{BND} + E_{BC} + E_{ANG} + E_{DIH} + E_{NB} + E_{CUT} + E_{GBSA} \quad (1)$$

where the terms have the following meaning:

E_{SS}—the solvent–solvent nonbonded energy,

E_{SX}—the solvent–solute nonbonded energy (and solvent cap restraint, if any),

E_{XX}—the solute–solute (intersolute) nonbonded energy,

E_{BND}—the bond stretching energy for the solutes,

E_{BC}—the energy for the interatomic harmonic constraints,

E_{ANG}—the angle bending energy for the solutes,

E_{DIH}—the torsional energy for the solutes,

E_{NB}—the >1,3 intramolecular non-bonded energy for the solutes,

E_{CUT}—the cutoff correction for the Lennard–Jones interactions neglected beyond the cutoff for nonaqueous solvents,

E_{GBSA}—optional GB/SA free energy of solvation.

Free energy changes, if desired, are computed with statistical perturbation theory in a windowing format with double-wide sampling. ΔG (Gibbs) and ΔA (Helmholtz) are computed for NPT and NVT simulations, respectively. FEP calculations are not implemented for GB/SA solvation. Otherwise, the expression for ΔG is

$$\Delta G = G_j - G_i = -kT \ln \langle \exp((E_j - E_i)/kT) \rangle_i \quad (2)$$

where the perturbation is from the reference system i to the perturbed system j. The mutation of A to B involves a scaling where a coupling parameter λ is used such that for λ = 0 the system is A and for λ = 1 it is B. Then for any geometrical variable or potential function parameters $Y(q, \sigma, \varepsilon)$,

$$Y_i = \lambda Y_B + (1 - \lambda) Y_A \tag{3}$$

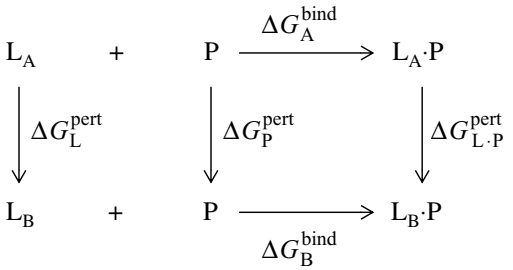
is used to scale the system from A to B. Thus, the simulation is run at a reference value λ corresponding to system *i* in eq 2. The program perturbs the system to two other values of λ corresponding to *j* and *k*, where typically $j < i < k$. This computation of two free energy increments in one simulation has been termed "double-wide sampling" (J. Chem. Phys. **1985**, *83*, 3050).

The number of free energy increments that need to be computed depends on the details of the overall perturbation, the solvent and T, P conditions. Many examples are available in the references given on page 3. For the conversion of ethane to methanol in water using the BOSS program, 5 increments were sufficient to give excellent precision (J. Chem. Phys. **1985**, *83*, 3050), while the annihilation of adenine or guanine in chloroform required 30–40 increments necessitating 15–20 simulations with double-wide sampling (Tetrahedron, **1991**, *47*, 2491). In a MCPRO simulation of a FK506 binding protein inhibitor in water, 13 simulations were used to remove a phenyl substituent.

The perturbations can also involve a reaction coordinate such as a dihedral angle or intermolecular distance. In this case there might be no change in the atom types (and potential function parameters).

2.3 Relative Binding Affinities

The calculation of the relative binding affinities of two ligands, L_A and L_B, can be represented as a thermodynamic cycle:



where ΔG_{A}^{bind} and ΔG_{B}^{bind} are the free energies of binding of the two ligands, ΔG_{L}^{pert} is the free energy of perturbation of L_A into L_B, ΔG_{P}^{pert} is the free energy of perturbation of the protein into itself (which is of course zero), and $\Delta G_{L \cdot P}^{pert}$ is the free energy of perturbation of the complex with ligand A into the complex with ligand B.

The relative binding free energy is the difference between the two free energies of binding, and can be expressed in terms of the thermodynamic cycle as follows:

$$\Delta\Delta G_{A \rightarrow B}^{bind} = \Delta G_{B}^{bind} - \Delta G_{A}^{bind} = \Delta G_{L \cdot P}^{pert} - \Delta G_{P}^{pert} - \Delta G_{L}^{pert} = \Delta G_{L \cdot P}^{pert} - \Delta G_{L}^{pert} \quad (4)$$

The free-energy perturbation calculations required involve perturbing ligand A to ligand B in the absence of the receptor (the “free” calculation) and in the presence of the receptor (the “bound” calculation).

Creating a Model System

To run an MCPRO+ job, you must first set up a model system. MCPRO+ provides an easy-to-use graphical interface for setting up the model system. The only input required is the set of structures you want to use, in Maestro-formatted files. You can also generate a Z-matrix from the command line with the utility `mcpro_zmat`, which is described in Chapter 11. This chapter describes the use of the MCPRO+ panel in Maestro to set up a model system.

The structural input must consist of properly prepared, 3D all-atom structures. For proteins, we strongly recommend that you use the Protein Preparation Wizard panel, which you can open from the Workflows menu in the Maestro main window. For more information, see the *Protein Preparation Guide*. For ligands, we recommend the use of LigPrep, which produces minimized 3D all-atom structures. For more information, see the *LigPrep User Manual*.

In addition to proper structural preparation, proteins must meet the following requirements for use with MCPRO+.

- Non-native residues may not be included in the chopped protein (unless they are added to the `pepz` database).
- Residue names have priority over the structure in neutralization and Z-matrix creation. It is therefore necessary for the structure to match the residue names.

The process of model system creation involves the following steps, for a protein-ligand complex:

1. Generate atomtyping parameters for the ligand species for the OPLS2001 or the OPLS2005 force field.
2. Generate the ligand Z-matrix with the `autozmat` program, after analyzing geometry differences (if any) between the initial and final structures to determine geometry variations to include.
3. Merge the atomtyping information for the ligand into the ligand Z-matrix and write out the Maestro formatted structure file with the same atom numbering as in the Z-matrix file. This creates the ligand model system.
4. Chop the protein and keep only residues within a given radius of the ligand. To minimize the number of disconnected chains, residues are included if they are part of a gap of 3 residues or less and isolated chains of 2 residues or less are removed.

5. Cap the "loose" ends of the chopped protein with ACE and NMA groups.
6. Neutralize the chopped and capped protein-ligand system. This is done in a different manner from that normally used with MCPRO, where all residues distant from the ligand are neutralized and residues are ionized to preserve system neutrality. In MCPRO+, residues are classified into different shells and then applicable residues (LYS, ARG, GLU, ASP, GLH, HIP, and so on) in the outermost shell from the ligand are ionized or neutralized to obtain a neutral formal charge for the system. If the charge could not be neutralized using only residues in the outermost shell, neutralization of residues in the second outermost shell is attempted.
7. Remove cofactors or metal ions from the protein structure and generate a Z-matrix for each of these species following the analogous procedures to steps 1–3.
8. Generate the protein Z-matrix with `pepz`.
9. Merge the protein and cofactor or metal Z-matrices, and write the corresponding Maestro file.
10. Merge the protein model system and ligand model system to form the protein-ligand complex model system.

For proteins only or for ligands only, the relevant parts of this list of tasks is performed. These tasks are all performed automatically when you use the MCPRO+ panel for model system creation.

In the MCPRO+ panel, the degrees of freedom to be varied in the sampling are also defined during creation of the model system creation, with the exception of explicit waters included as part of the protein. Solvent water molecules are not part of the model system, but are added when the job is run.

Setting up the model system is done in the Create Model System tab of the MCPRO+ panel. The Create Model System tab provides a wizard-like interface that takes you through the stages of setting up a model system. For new systems that contain proteins, there are five steps. For ligand-only systems, two of these steps are not present.

Each step contains the main controls for the step, at the top, then a Display Options section, then a Guide at the foot of the panel. The Guide consists of a button for each step, connected by arrows. The buttons for steps that are not yet available are dimmed. The background of the step button for the current step is white; the others are gray. The Display Options section contains options for displaying the various structures belonging to the model in the Workspace.

The main controls for each step, and the common Display Options section, are described below.

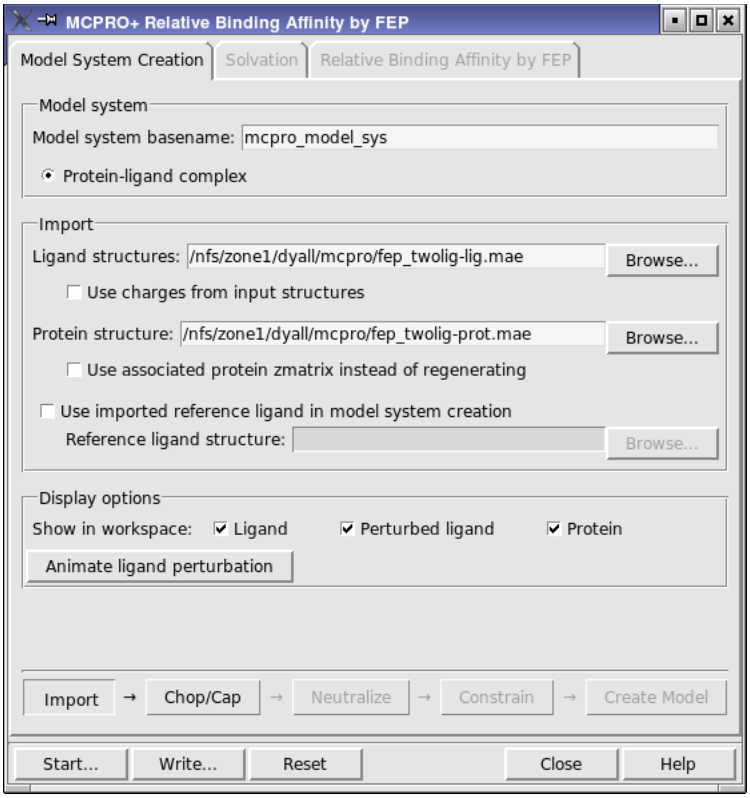


Figure 3.1. The Import step in the Model System Creation tab.

3.1 Importing the Structures

In the Import step, you define the type of system and import the structures for the calculation. There are two sections in the main part of the step, the Model system section and the Import section.

3.1.1 Defining the Model System Type

In the Model system section, you choose the basic types of structures to be included in the system, and set a base name for the model system. The base name is used to name the files that are created to store information for the model system.

There are three possible options for the system type. These options vary according to the type of calculation, and can include one or more of the following:

- Protein-ligand complex
- Ligands only
- Protein only

When you choose a system option that does not contain a protein, the Chop/Cap and Neutralize steps are removed from the Guide, since they are no longer relevant.

3.1.2 Importing the Structures

In the Import section, you import the structures for the system. The structures must be valid structures in Maestro or SD format.

The receptor and ligand structures that you select must be 3D all-atom structures with correct charges and bond orders. To prepare the receptor structure, you can use the Protein Preparation Wizard, which is described in Chapter 2 of the *Protein Preparation Guide*. To prepare the ligand structures you can use LigPrep—see the *LigPrep User Manual* for details.

The controls available depend on the choice of the system type in the Model system section. The possible sets of controls with their descriptions are given below:

Ligand structures

To specify the ligand structures, enter the name of the ligand structure file in the text box, or click Browse and navigate to the file. The file name is displayed in the text box. The number of ligands to be imported depends on the type of job.

- FEP and relative binding affinity—the input ligand structure file must have two structures with the same topology and atom numbering. Any other structures are ignored.
- Linear response—as many ligands as you like, but all must have the same total formal charge.
- MC and Minimization—only one ligand. Any other structures in the file are ignored.

Ligands must be in the correct frame of reference (i.e. of the receptor) and must have no more than 200 atoms.

The ligand is imported and displayed in the Workspace when you have entered the file name or clicked OK in the file chooser.

You can also select Use charges from input structures to make use of the partial charges stored with the ligands, rather than generate them from the force field.

Protein structure

To specify the protein, enter the name of the protein structure file in the text box, or click Browse and navigate to the file. The file name is displayed in the text box. The protein structure must not include the ligand structures specified as input, but can include cofactors, metal ions, and so on. The protein is imported and displayed in the Workspace when you have entered the file name or clicked OK in the file chooser.

If the protein was used previously for a model system, you can select Use associated protein zmatrix to skip over the protein preparation steps. Selecting this option removes the Chop/Cap and Neutralize steps from the Guide.

For protein-ligand complexes, you can choose to import a reference ligand that will be used in subsequent steps, for protein truncation and for defining the degrees of freedom to sample. To do so, select Use imported reference ligand in model system creation, then enter the name of the ligand structure file in the text box, or click Browse and navigate to the file. The file name is displayed in the text box.

For protein-only calculations, you *must* import a reference ligand to be used in subsequent steps, for protein truncation, neutralization, and defining the degrees of freedom to sample. To do so, enter the name of the ligand structure file in the Reference ligand structure text box, or click Browse and navigate to the file. The file name is displayed in the text box. Only the first ligand structure in the file is used.

When you have made your choices, the button for the next step becomes available in the Guide, and you can click it to proceed.

3.2 Truncating and Capping the Protein

Truncating the protein is natural within a Monte Carlo simulation that uses finite nonbonded cutoff distances where a portion of the system is treated rigidly. Residues beyond the nonbonded cutoff distance of the flexible region have a constant contribution to the system energy and can therefore be neglected, resulting in a significant speed-up in the simulation.

Keeping more atoms in the system than are needed is likely to degrade the results because the degrees of freedom that matter are not being sampled as much. Therefore, the system should be truncated to within some distance of the actively sampled region. Reducing the size of the system also reduces the number of water molecules needed if a solvent cap is used. In addition, MCPRO has a limit of 9999 atoms for protein-ligand complexes.

The purpose of the Chop/Cap step is to truncate the protein and cap the truncated parts with NMA or ACE groups. This step is only available when a protein is being used. There is only one section for making settings, Chop/Cap protein.

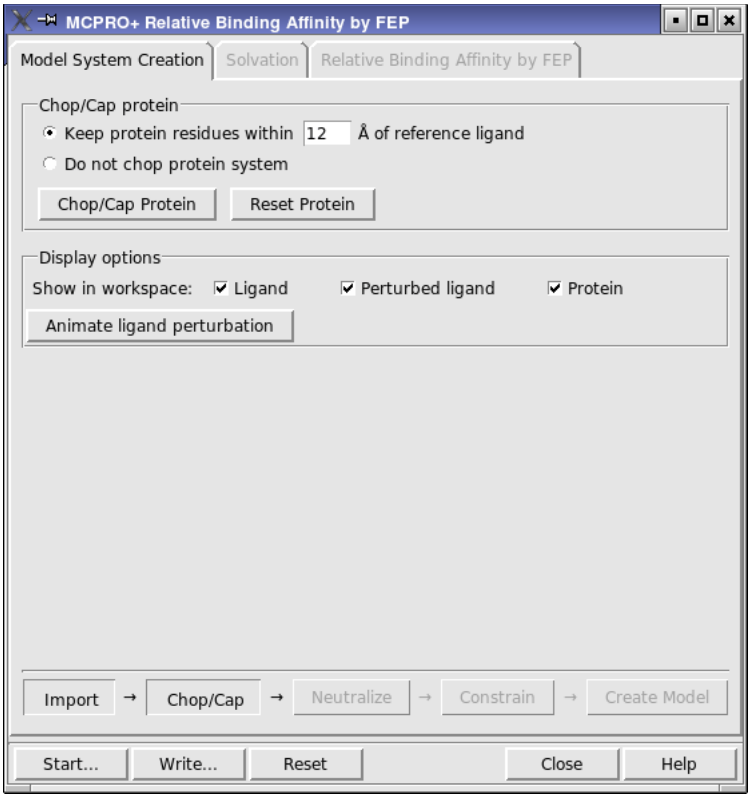


Figure 3.2. The Chop/Cap step in the Model System Creation tab.

To truncate and cap the protein:

1. Select Keep protein residues within *N* Å of reference ligand and enter a distance in the text box.

   All residues that are within the specified distance of any atom in the ligand are kept, and the residues at the cut are capped with NMA or ACE as appropriate. The reference ligand is used for this purpose if one has been provided, otherwise the first ligand is used. A smaller protein model also means that the number of solvent water molecules added to the system will be smaller.

   Residues inside the specified distance are colored yellow, residues outside this distance are colored blue. This allows you to check whether the truncation distance is appropriate, and adjust it as necessary. When you enter a new value, the coloring is updated.

2. Click Chop/Cap Protein

   The truncation job is started, and a message is displayed in the lower part of the panel while the job is running. After the job finishes, the next step, Neutralize, becomes available. The capping groups are now displayed in blue.

If you decide that you kept too much or too little of the protein, you can click Reset Protein to discard the results of the truncation and return to the original protein. You can then change the distance and rerun the truncation.

As discussed above, chain integrity is preserved as far as possible, so small segments (1 or 2 residues) with gaps of 3 or less residues are joined in; further away they are removed.

If you do not want to truncate the protein, select Do not chop protein system. The next step, Neutralize, then becomes available.

3.3 Neutralizing the System

The purpose of the Neutralize step is to ensure overall electrical neutrality of the protein or protein-ligand complex. This is important to obtain good results. Neutralization is performed by adding or removing protons from ionizable residues that are distant from the ligand. This step is only available when a protein is being used.

There is one section in the step, Neutralize system, which has two options for neutralizing the protein, and action buttons to perform or undo the neutralization task. The options are:

Neutralize to first ligand charge

Select this option to neutralize the complex by adding or removing protons from the protein so that the protein charge balances the ligand charge, and the charge on the complex is zero.

Target protein charge

Select this option to add or remove protons from the protein so that the protein charge is equal to the value given in the text box.

When you have decided how to neutralize the protein, click Neutralize Protein to perform the neutralization. A message is displayed in the lower part of the panel while the neutralization job is running. When the job finishes, the residues that were deprotonated are colored red, those that were protonated are colored yellow, and the remaining residues are colored gray.

You must click the Neutralize Protein button to neutralize the protein and proceed to the next step. After the protein is neutralized, the next step becomes available.

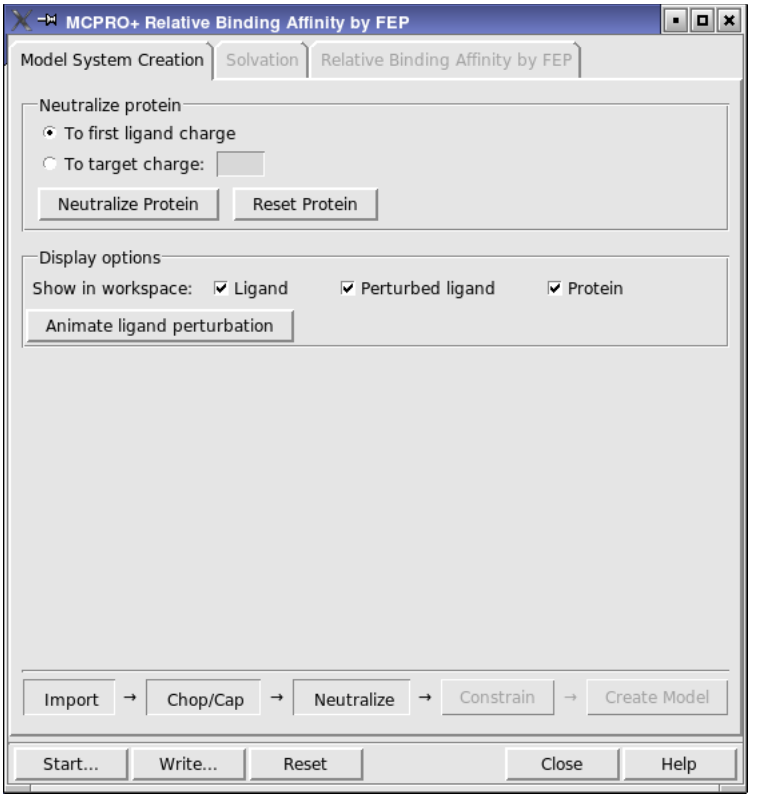


Figure 3.3. The Neutralize step in the Model System Creation tab.

If you want to change the neutralization, click Reset Protein to reset the protein to its original state, before addition or removal of protons.

3.4 Defining the Degrees of Freedom

In the Constrain step, you choose which parts of the protein are treated flexibly, and whether the ligands or cofactors are kept rigid or allowed to be flexible. The choices you make in this section limit the choice of minimization method (see Chapter 5). You can also choose the force field for the simulation. These controls are in the Constrain molecules section.

There are three possible sets of options, and the options available depend on the system.

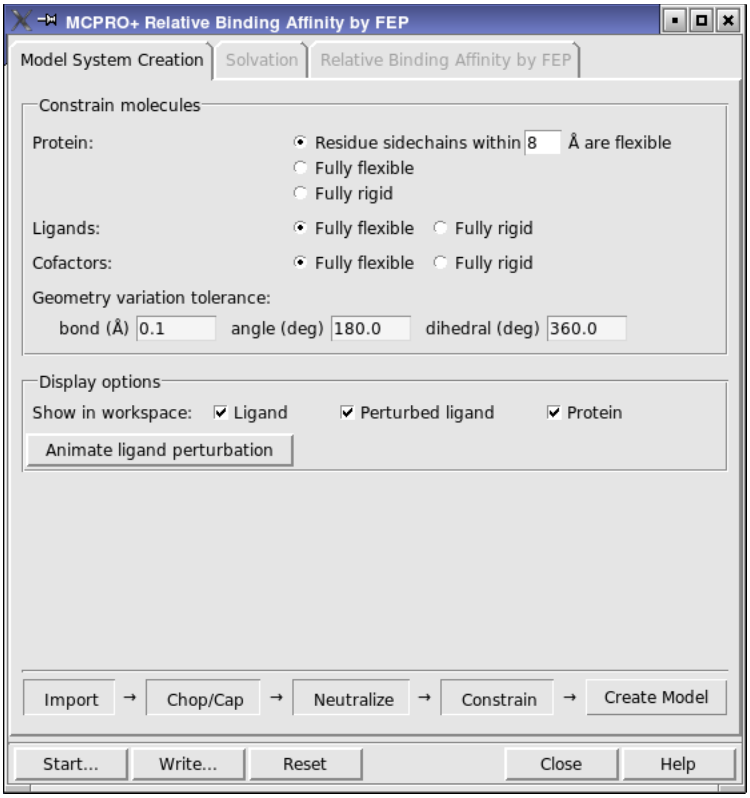


Figure 3.4. The Constrain step in the Model System Creation tab.

There are three options for constraining the protein:

- Residue side chains within *N* Å are flexible—Atoms in the side chains of residues that are within the specified distance of the ligand or reference ligand are allowed to move. All other atoms are held fixed, including the backbone atoms for residues whose side chains are flexible.
- Fully flexible—All atoms in the protein are allowed to move.
- Fully rigid—No atoms in the protein are allowed to move.

For the ligand and the cofactors, there are only two options: Fully flexible, for which all atoms can move, and Fully rigid, for which all atoms are kept fixed.

The regions that you select are marked in the Workspace. Flexible regions of protein are colored yellow, fixed regions are colored blue.

For the protein, only the bond angles and dihedral angles are sampled, but for the ligand, bond lengths are sampled as well.

For free-energy perturbation calculations and relative binding affinity predictions, the Geometry variation tolerance text boxes are displayed in this section. You can use these to set tolerances for the geometric parameters (bond lengths, angles, and dihedrals) that are considered to be the same in the two ligands. Parameters that differ by less than the specified tolerance are considered to be the same. Parameters that are considered different are driven in the FEP steps, so it is important to keep the number of these parameters to a minimum.

This step completes the selection of options.

3.5 Creating the Model

When you click Create Model, a job is run to set up the files for the model system, and the final step is displayed when the job finishes.

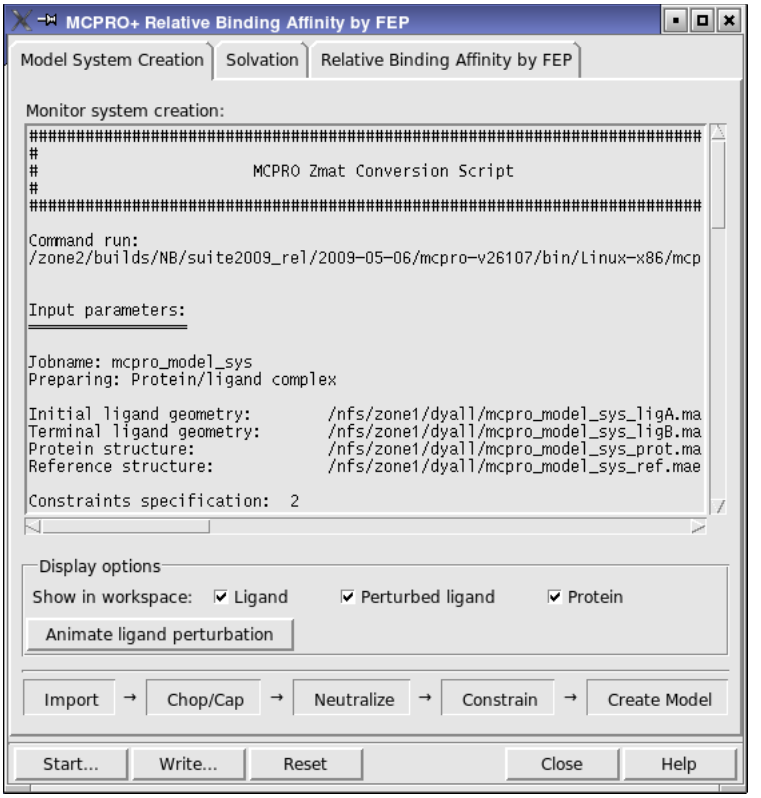


Figure 3.5. The Create Model step in the Model System Creation tab.

The Create Model "step" presents the results of the creation of the model system files. No action is required. You can examine the log file for the job that writes the files to ensure that the system is set up correctly and that there were no errors.

3.6 Display Options

The Display Options section, which is present in each step, contains options for the display of each of the structures in the model system. The controls depend on the model system and the calculation type. For most models, there is a list of options for each structure. For FEP calculations, there is a Ligand Perturbation button, which animates the perturbation in the Workspace.

3.7 Files Generated

Creation of a model system generates a Z-matrix file and a Maestro file for each part of the system: the protein, the ligands, and the complex:

basename`-comp.mae`	Maestro format component of the model system for the complex.
basename`-comp.zmat`	MCPRO$^+$ Z-matrix format component of the model system for the complex.
basename`-prot.mae`	Maestro format component of the model system for the protein, including cofactors.
basename`-prot.zmat`	MCPRO$^+$ Z-matrix format component of the model system for the protein, including cofactors.
basename`-lig.mae`	Maestro format component of the model system for the ligands
basename`-lig.zmat`	MCPRO$^+$ Z-matrix format component of the model system for the ligands
basename`.log`	Log file from model system creation

All three model systems are present if generation was done for a complex. For a protein only or for a ligand only, there is only one model system created, with the relevant name from the list above.

Solvation Models

MCPRO+ supports both continuum solvation and explicit solvent models. The models can be selected in the Solvation tab. There are up to five choices for a solvation model, depending on the type of calculation and the model system. Some choices are further restricted by choices made in other tabs. A summary of the available models is given in Table 4.1.

Table 4.1. Solvent model availability. P stands for protein and L stands for ligand.

Solvent Model	Minimization	FEP	Relative Binding Affinity	MC Sampling	Linear Response
Continuum with specified dielectric	P,L	P, L	P, L	P,L	
Continuum, GB/SA water	P,L			P,L	
Explicit water, periodic		L		L	
Explicit water, cluster		P,L	P,L	P,L	P,L
Gas Phase	P,L	P,L	P,L	P,L	

In addition to running calculations in a solvent, you can run a calculation in the gas phase, by selecting Gas phase in the Solvation tab.

4.1 Continuum Solvation Models

Two continuum solvation options are available:

Constant dielectric of x

This option allows you to specify a general solvent using a continuum model. You can specify the dielectric constant in the text box. If you are minimizing a structure with the steepest descent or conjugate gradient method, you can choose to use a distance-dependent dielectric. The minimization method is chosen in the Minimization tab.

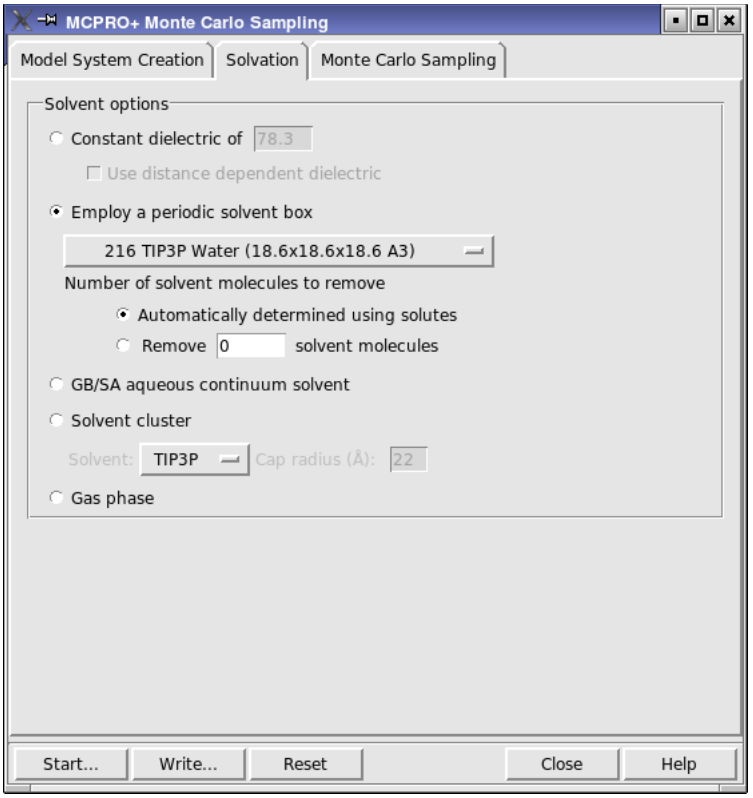


Figure 4.1. The Solvation tab.

GB/SA aqueous continuum solvent

This option allows you to run the simulation using the GB/SA model for aqueous continuum solvation. For minimization calculations, this solvation method is available with all except the conjugate gradient method.

This option is not available for FEP or linear response calculations.

4.2 Explicit Solvent Models

MCPRO simulations can be run with explicit water molecules, using the TIP3P or TIP4P model for the water molecules, and using periodic boundary conditions or a solvent cluster. In both cases, pre-equilibrated sets of water molecules are used, and solvent molecules are removed to make room for the solute.

Use of a solvent cluster is the only solvent model available for linear response calculations.

To run a simulation with periodic boundary conditions:

1. Select Employ a periodic solvent box.

   The controls under this option become available.

2. Choose the solvent model (TIP3P or TIP4P) and the box dimensions from the option menu.

   Each item on the menu corresponds to a combination of solvent and box size, which determines the number of water molecules in the box before the solute (protein, ligand, or complex) is added to the box.

3. Select an option for removal of solvent molecules.

   Solvent molecules must be removed to accommodate the solute, and are chosen for removal in order of their interaction energy with the solute, from highest to lowest, i.e. high-energy molecules are removed first. In the automatic removal, one water molecule is removed per solute heavy atom.

Note: If you choose to specify the number of solvent molecules to remove, you should make an accurate estimate, otherwise the job could fail or the results of the calculation might be erroneous.

To run a simulation in a solvent cluster:

1. Select Solvent cluster.

   The controls under this option become available.

2. Choose the solvent model from the Solvent option menu.

   The two models available are TIP3P and TIP4P.

3. Enter a radius in the Cap radius text box.

The radius is measured from the centroid of the ligand (either the reference ligand, if one was selected, or the first ligand specified for the simulation). The solvent molecules are added from pre-equilibrated caps of the given radius, and solvent molecules within a given distance of solute atoms are removed. The solvent cap is marked in the Workspace by a translucent blue sphere, so you can see whether any part of the structure is not encompassed by solvent.

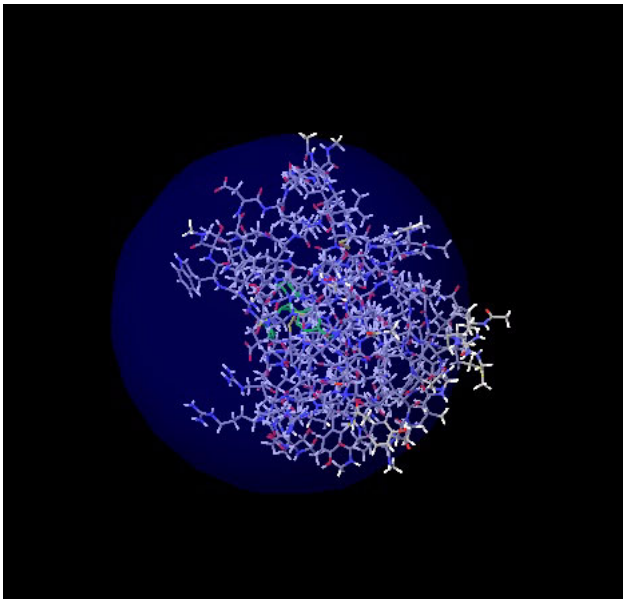


Figure 4.2. The solvent sphere marked in the Workspace.

Minimization Calculations

MCPRO+ can perform minimization calculations with the selected force field. It does so by calculating gradients by finite difference, rather than analytically. The performance in programs that do use analytic gradients is usually better, so these programs are generally preferred for minimizations.

To set up a minimization calculation with MCPRO+, choose Minimization from the MCPRO+ submenu of the Applications menu. In the Minimization tab, you choose the optimizer for the minimization process and set the number of cycles for the minimization. The choice of optimizer also influences the availability of some solvation options—see Chapter 4.

You can set the number of cycles to use in the Number of optimization cycles text box, and then choose the minimization method from the Method option menu. The available methods are:

- CONJUG—Conjugate gradient method. Only available when the protein, ligand, or both, are fully flexible.
- SD—Steepest descent method. Only available when the protein, ligand, or both, are fully flexible.
- SIMPLEX—Downhill simplex method. Only available when the protein, ligand, or both, are only partly flexible.
- FLEPOW—Davidson-Fletcher-Powell method. Only available when the protein, ligand, or both, are only partly flexible.

Completely flexible systems are minimized in Cartesian coordinates, so are much faster than partly flexible systems, which use internal coordinates for minimization. Internal coordinate minimization has a limited number of degrees of freedom that can be used, so it is not recommended for proteins, but is best to use for ligands.

The properties that are reported in the text output and the output Maestro file are listed in Table 5.1. The last five properties are only reported if GB/SA solvation is used.

Table 5.1. Properties generated in a minimization

Property	Description
Bond Energy (kcal/mol)	Bond stretching energy for the solutes.
Angle Energy (kcal/mol)	Angle bending energy for the solutes.
Torsion Energy (kcal/mol)	Torsional energy for the solutes.
Intrasolute Non-Bonded Energy (kcal/mol)	Intra-solute Coulomb + van der Waals energy for the solutes.
Intersolute Non-Bonded Energy (kcal/mol)	Inter-solute Coulomb + van der Waals energy for the solutes. Only computed for protein-ligand complexes.
Intersolute Lennard-Jones Energy (kcal/mol)	Inter-solute van der Waals energy. Only computed for protein-ligand complexes.
Intersolute Coulomb Energy (kcal/mol)	Inter-solute Coulomb energy. Only computed for protein-ligand complexes.
Bond Constraints Energy (kcal/mol)	Energy within constrained bonds. (Bond constraints cannot be set up from Maestro.)
Total Energy Without GB/SA (kcal/mol)	Total system energy without including GB/SA contributions.
GB/SA Component: G Polarization (kcal/mol)	GBSA polarization energy component.
GB/SA Component: G SASA	GBSA solvent accessible surface area component.
GB/SA Contribution (kcal/mol)	Total GBSA contribution to the system energy.
Total Energy + GB/SA (kcal/mol)	Total system energy including GBSA contribution.

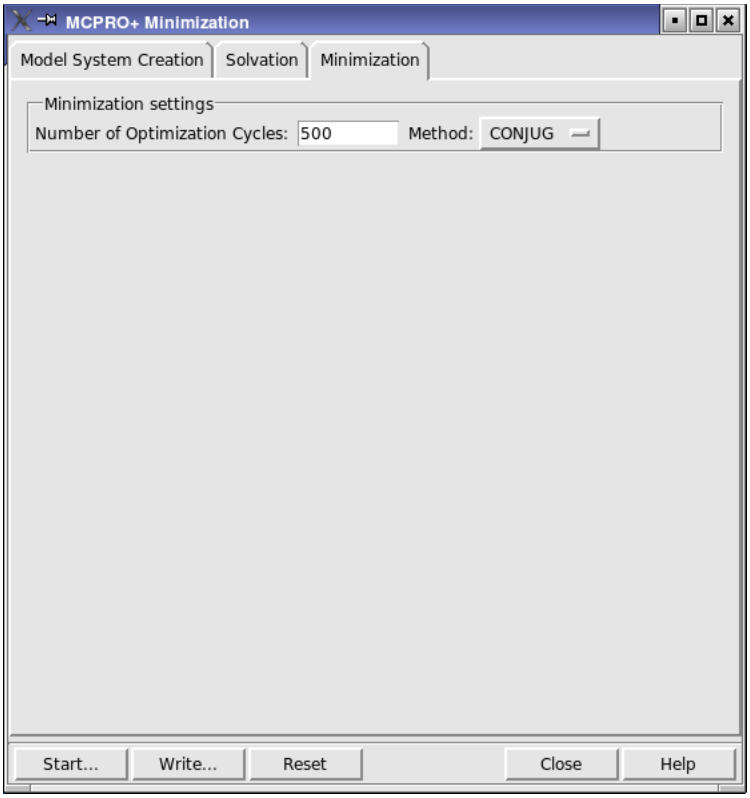


Figure 5.1. The Minimization tab.

Monte Carlo Simulations

MCPRO+ provides an interface for general-purpose Monte Carlo simulations. Monte-Carlo simulations form the basis for most MCPRO+ job types, so much of the information recorded here applies also to other job types, such as free-energy perturbation and binding affinity prediction.

To set up a Monte Carlo simulation, choose Monte Carlo Simulation from the MCPRO+ submenu of the Applications menu.

6.1 Setting Up the Simulation

Once you have set up a model system and chosen a solvation method, you can use the Monte Carlo tab to set up the basic Monte Carlo simulation parameters. This tab has two sections.

In the Monte Carlo settings section, you set the pressure and temperature for the ensemble, and specify the block size. Each block is run as a separate MCPRO+ calculation, but the blocks are run sequentially to maintain the Markov chain. If you are using a periodic solvent box, the Ensemble option menu is also present, from you can also choose an ensemble. The choices are NPT and NVT.

In the MC run section you select the stages of the simulation and specify parameters that control the length of the simulation. The choices are:

Solvent-only equilibration for *N* configurations

Select this option to equilibrate the solvent in the presence of the solute (protein, ligand, or complex), and specify the number of configurations for the equilibration in the text box. Only the solvent is sampled.

System equilibration for *N* configurations

Select this option to equilibrate the entire system, and specify the number of configurations for the equilibration in the text box. This step is performed after any solvent equilibration. All degrees of freedom are sampled.

System averaging for *N* configurations

Select this option to perform sampling of the system to obtain the final results, and specify the desired number of configurations in the text box.

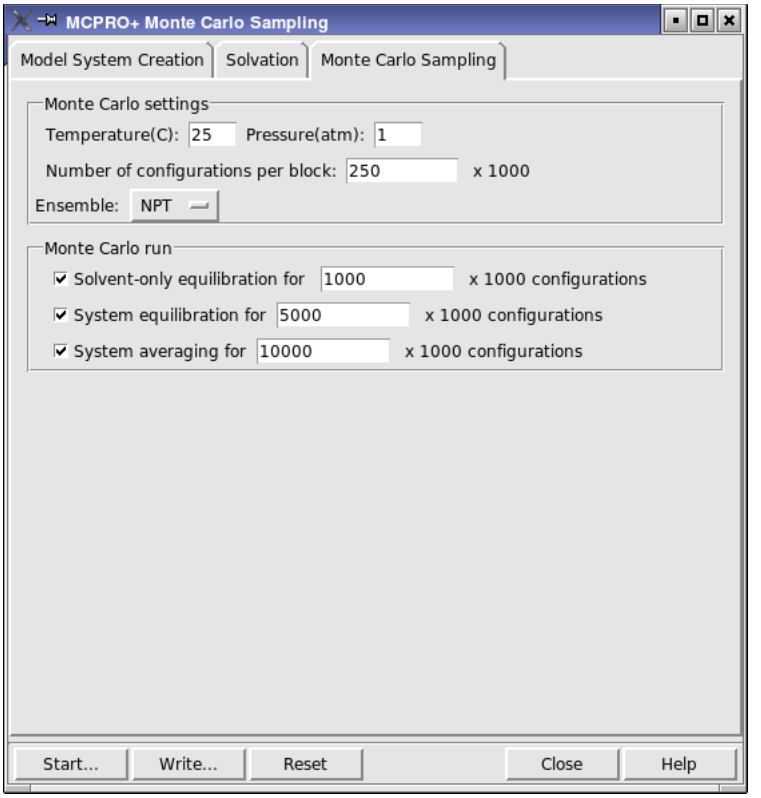


Figure 6.1. The Monte Carlo Sampling tab.

When you have made selections, click Start. The Start dialog box opens, in which you can make job settings and submit the job to a host for execution.

6.2 Output Files

Output files are generated for each run of the MCPRO$^+$ executable. This means that output files are generated for each block in the simulation. The output files for a block include a compressed Maestro file, *name*`.maegz`, and a text output file, *name*`.out`, where *name* is generated by the driver.

The output Maestro file for each block contains the last accepted structure, including waters, and a set of properties that include the static energies of the final configuration as well as the averages. The text output file includes the properties for each block completed to that point. The properties that are reported for each block are the per-block properties and the per-block

averaged properties listed in Table 6.1. The per-block averaged properties are "running averages" over the blocks completed to that point.

The output files (Maestro and text) for the entire simulation include all the properties listed in Table 6.1. The per-block properties reported in the final Maestro file are the values for the last block, and the structure is the last accepted structure in the last block. The text output file contains properties for all the blocks in the simulation.

You can read the `.out` files into the Analysis panel, and display plots of the various properties as a function of number of configurations. Since each `.out` file for a block contains results for all the blocks run to that point, you can use them to plot the progress of the simulation. The Analysis panel is described in Section 7.4 on page 43.

Table 6.1. Properties generated in a Monte Carlo simulation

Property	Description
Miscellaneous	
Solute1 Percent Moves Accepted	Percent of moves accepted for the first solute. Reasonable values typically range between 20% and 60% with a target of around 40%
Solute2 Percent Moves Accepted	Percent of moves accepted for the second solute. For protein-ligand calculations this is the ligand. Reasonable values typically range between 20% and 60% with a target of around 40%
Number of Configurations Sampled	Total number of configurations examined in the current sampling. This typically includes many blocks of MC sampling.
Per-block properties, averaged over configurations in the block	
Total System Energy (kcal/mol)	Total system energy.
Solute-Solvent Energy (kcal/mol)	Non-bonded energy between all solutes and solvent.
Solute-Solute Energy (kcal/mol)	Non-bonded energy between solvent molecules.
Solute Bond Energy (kcal/mol)	Bond stretching energy for solutes.
Solute Angle Energy (kcal/mol)	Angle bending energy for solutes.
Solute Torsion Energy (kcal/mol)	Torsional energy for solutes.

Table 6.1. Properties generated in a Monte Carlo simulation

Property	Description
Solute Non-Bonded Energy (kcal/mol)	Non-bonded energy for solutes. Includes solvent-solute and solute-solute contributions.
Solute Bond Constraints Energy (kcal/mol)	Energy within constrained bonds. (Bond constraints cannot be set up from Maestro.)
Solvent-Solute*N* Coulomb Energy (kcal/mol)	Coulomb energy between solute *N* and solvent.
Solvent-Solute*N* LJ Energy (kcal/mol)	Van der Waals energy between solute *N* and solvent.
Solute-Solute Coulomb Energy (kcal/mol)	Coulomb energy between solutes.
Per-block averaged properties, averaged over blocks completed so far in the simulation	
Bond Energy (kcal/mol)	Bond stretching energy for the solutes.
Angle Energy (kcal/mol)	Angle bending energy for the solutes.
Torsion Energy (kcal/mol)	Torsional energy for the solutes.
Intrasolute Non-Bonded Energy (kcal/mol)	Intra-solute Coulomb + van der Waals energy for the solutes.
Intersolute Non-Bonded Energy (kcal/mol)	Inter-solute Coulomb + van der Waals energy for the solutes. Only computed for protein-ligand complexes.
Intersolute Lennard-Jones Energy (kcal/mol)	Inter-solute van der Waals energy. Only computed for protein-ligand complexes.
Intersolute Coulomb Energy (kcal/mol)	Inter-solute Coulomb energy. Only computed for protein-ligand complexes.
Bond Constraints Energy (kcal/mol)	Energy within constrained bonds. (Bond constraints cannot be set up from Maestro.)
Total Energy Without GB/SA (kcal/mol)	Total system energy without including GB/SA contributions.
GB/SA Component: G Polarization (kcal/mol)	GBSA polarization energy component.
GB/SA Component: G SASA	GBSA solvent accessible surface area component.
GB/SA Contribution (kcal/mol)	Total GBSA contribution to the system energy.

Table 6.1. Properties generated in a Monte Carlo simulation

Property	Description
Total Energy + GB/SA (kcal/mol)	Total system energy including GBSA contribution.
Standard deviations of per-block averaged properties	
stdev Total System Energy (kcal/mol)	Standard deviation of total system energy.
stdev Solvent-Solute*N* Coulomb Energy (kcal/mol)	Standard deviation of Coulomb energy between solute *N* and solvent.
stdev Solvent-Solute*N* LJ Energy (kcal/mol)	Standard deviation of van der Waals energy between solute *N* and solvent.
stdev Solute-Solute Coulomb Energy (kcal/mol)	Standard deviation of Coulomb energy between solutes.
Averaged properties over all blocks	
average Total System Energy (kcal/mol)	Total system energy, averaged over all blocks.
average Solvent-Solute*N* Coulomb Energy (kcal/mol)	Coulomb energy between solute *N* and solvent, averaged over all blocks.
average Solvent-Solute*N* LJ Energy (kcal/mol)	Van der Waals energy between solute *N* and solvent, averaged over all blocks.
average Solute-Solute Coulomb Energy (kcal/mol)	Coulomb energy between solutes, averaged over all blocks.

Free-Energy Difference Calculations

MCPRO+ can perform free-energy perturbation (FEP) calculations on ligands, proteins and complexes. The background of FEP calculations is described in Section 2.2 on page 6. FEP calculations are the basis of relative binding affinity predictions, which are treated in the next chapter.

As well as setting up a model system, there are other requirements on the structures you use for FEP calculations. These requirements and instructions on how to satisfy them are given in the next section. The following section describes how to set up a FEP calculation from Maestro.

MCPRO+ also provides graphical analysis tools for the FEP calculations, in which various quantities are plotted as a function of the perturbation parameter. These tools are described in Section 7.4.

7.1 Setting Up Structures for Free-Energy Perturbation

Free-energy perturbation simulations are not intended for major structural changes. When you set up structures for perturbation, you should make only small changes, such as replacement of a hydrogen atom with a methyl group; conversion of a methyl group to an amino group, a hydroxyl group, or a doubly-bonded oxygen; or conversion of CH_2 to NH or O. You should also avoid making many changes to convert one structure to another. Large changes require much longer simulation times, and might not yield the desired accuracy.

The changes in the structure need not be in the functional groups, they can be geometric. MCPRO+ looks for changes in the geometry as well as changes in the functional groups when it determines which parameters to drive in the perturbation process. If you use optimized structures for the initial and final state, MCPRO+ might include a large number of geometric parameters in the perturbation. If the geometric changes are not too large, you can loosen the thresholds for detecting geometric variation in the Constrain step of the Create Model System panel, so that small changes are not driven in the perturbation.

The structures that you use for an FEP calculation must have the same atom ordering (by atom number) in the structures at either end of the perturbation, so that the proper associations between the structures can be made. If the atom numbering and the connectivity do not match, the results are likely to be erroneous.

In order to run an FEP calculation to convert one structure into another structure with different numbers of atoms, dummy atoms must be added to the smaller structure that correspond to atoms in the larger structure. When preparing structures for FEP, it is therefore a good idea to use the larger structure to prepare the smaller structure, by converting atoms into dummy atoms.

Note: If you want to use partial charges obtained from some other program, such as Jaguar, you must prepare the structures with the dummy atoms first, before running the calculation to produce the partial charges. This is necessary to ensure the correct atom numbering.

As an example of structure conversion, if you wanted to convert a methyl group into a hydrogen, you would create the structure with the hydrogen from the structure with the methyl by converting the methyl hydrogens to dummy atoms first, then converting the methyl carbon to a hydrogen atom.

You can make these changes with the tools in the Build panel, as follows:

1. Make a duplicate of the larger structure (Duplicate on the Entry menu in the Project Table).

   This structure will become the smaller structure, so you may want to rename it.

2. Display the duplicate structure in the Workspace.

3. Open the Build panel, by clicking its toolbar button:

4. In the Atom Properties tab, choose Atom Type (MacroModel) from the Properties option menu.

5. Select Special dummy atom type (FEP) from the list (symbol Du, number 61).

6. Click on the atoms that you want to convert to dummy atoms.

   Always work from the terminal atoms inwards, and change multiple bonds to single bonds.

7. Select the appropriate type for any other kind of atom that you want to change and click on the atoms in the Workspace.

   You might have to cycle between converting to dummy atoms and converting to real atoms to achieve the result you want.

Other structural alterations, in which atoms are added and subtracted, are more difficult. You can convert one structure into the other using the tools in the Build panel, but you must be careful how you do the conversion. When you replace one fragment with another, Maestro first removes the old fragment, renumbers the atoms, then adds the new fragment. If the fragment you remove has the highest atom numbers in the structure, there is no problem, but if it does not, then you must convert atom types for the existing atoms, then add new atoms. Likewise, if you delete atoms, you must ensure that the atoms you delete have the highest atom numbers. Since Maestro does not provide user-controlled atom numbering, this approach is likely to fail to produce the desired numbering scheme.

To ensure that you have a common atom numbering scheme, you will probably have to build a structure that contains the atoms from both structures, duplicate this "supermolecule", then convert the atom types in each structure to produce the beginning and ending structures with the appropriate dummy atoms.

The following example converts aspirin to tylenol. This is an exercise in structure conversion, and should *not* be used to run an FEP calculation. This type of conversion, with simultaneous changes in several polar groups and the presence of acids, would require a very large number of configurations to be sampled for acceptable accuracy.

To convert aspirin to tylenol, the acid group in the 2 position of aspirin must be replaced with a hydroxyl group in the 4 position, and the oxygen of the ester converted to an amide NH group. To create a structure that contains atoms in all the locations of both structures, tylenol is the better starting point. You will first add groups to the tylenol structure, then duplicate the structure and perform the atom type conversions.

1. Click the Import Structures button on the toolbar.

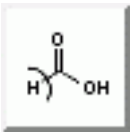
   The Import panel opens.

2. Navigate to the `$SCHRODINGER/mcpro-v`*version*`/tutorial` directory and select the file `tylenol.mae`.

3. Click Import.

   The tylenol structure is displayed in the Workspace.

4. In the Build panel, click the Acid (C->O) fragment button.

The acid fragment is selected for building.

5. Click the H atom that is ortho to the amide group in the tylenol structure. (Choose the one to the left.)

   The H atom is replaced with the acid group.

This molecule now has atoms in all the necessary locations for both structures, and they have the same numbering scheme. The atoms do not necessarily have the correct atom type, but that will be changed in the steps to follow.

6. Click the Open/Close Project Table button on the toolbar.

   The Project Table panel opens.

7. From the Entry menu, choose Duplicate.

   The entry is duplicated and selected (highlighted in yellow). Note that the original entry is still displayed in the Workspace (its In column is checked).

8. Click the Title column for the selected entry, and change the name to `aspirin`. (You can also change the Entry Name to aspirin.)

There are now two entries, one for each end of the FEP. In the steps to follow, you will convert atoms to dummy atoms in each structure, and retype any atoms that have the wrong atom type. The order in which the atoms are converted to dummy atoms is important: you must start from the terminal atoms and work in.

9. In the Atom Properties tab of the Build panel, choose Atom Type (MacroModel) from the Property option menu.

   A list of atom types is displayed in the center of the panel, and the Set atom type (MacroModel) option should be selected.

10. Scroll down until you see the Du atom type (number 61, described as "Special dummy atom type (FEP)"), and select this row.

11. Click on the H atom then on the O atom of the acid group.

    The atoms turn green and the bond length changes to 0.3 Å. This is the default bond length for a bond to a dummy atom, and is required because the force constant for a dummy atom stretch is large, and the energy terms from this stretch are included in the calculation.

For the carbonyl oxygen, we need to change the bond order first, and then change the atom type.

12. Right-click and hold on the carbonyl bond to open the Bond shortcut menu, and choose Single from the Order submenu.

    The bond changes to a single bond, and the two atoms of the bond are marked with yellow dots. (This is the "Workspace selection", and can be cleared by deselecting Set atom type (MacroModel) and clicking in the Workspace. It is not necessary to do this, but if you do, make sure that you reselect Set atom type (MacroModel) to continue with the exercise.)

13. Click on the O atom whose bond order you just changed.

    The atom turns green and the bond length changes to 0.3 Å.

14. Select the H1 atom type from the list (number 41).

15. Click on the carbon of the acid group.

    The carbon now becomes a hydrogen atom.

The structure has now been converted back to tylenol, but with dummy atoms in place for the FEP to aspirin.

Next, the aspirin structure must be converted.

16. Click the In column for the aspirin entry in the Project Table.

    The combined structure is displayed in the Workspace.

17. Ensure that Set atom type (MacroModel) is selected in the Build panel, and select the Du atom type.

18. Click on the phenolic hydrogen and the amide hydrogen.

    The atoms turn green and the bond length changes to 0.3 Å.

19. Select the H1 atom type from the list (number 41) and click on the phenolic oxygen.

    The oxygen is changed to a hydrogen.

20. Select the O3 atom type from the list (number 16) and click on the amide nitrogen.

    The nitrogen is changed to an oxygen.

The structure has now been converted back to aspirin, but with dummy atoms in place for the FEP to tylenol.

7.2 Setting Up a Free-Energy Perturbation Calculation

MCPRO+ FEP calculations require two ligand structures, which must be in the same Maestro file. These structures must be properly prepared, as described in the previous section.

Free-energy perturbation calculations are set up in the MCPRO+ Free Energy Difference by FEP panel. To open this panel, choose Free Energy Difference by FEP from the MCPRO+ submenu of the Applications menu.

After you have set up your model system and made solvent settings, you can set up the simulation parameters in the Free Energy Difference by FEP tab. This tab has three sections:

Monte Carlo settings

In this section you set the pressure and temperature for the ensemble, and specify the block size. Each block is run as a separate MCPRO+ calculation with the given number of configurations in the block, but the blocks are run sequentially to maintain the Markov chain.

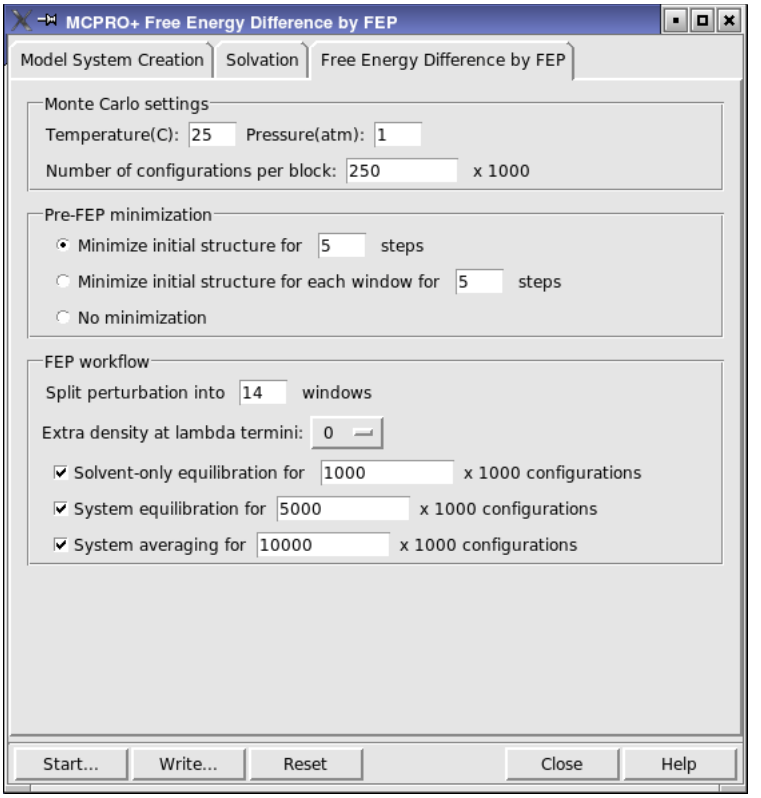


Figure 7.1. The Free Energy Difference by FEP tab.

Pre-FEP Minimization

In this section you can request preminimization of the two structures for a specified number of steps, prior to starting the FEP workflow or at the beginning of each step. The preminimization is performed on the entire solute, and is intended to relieve any strain resulting from model system creation. You can choose one of the following options, and enter the number of steps in the appropriate text box:

- Minimize initial structure for *N* steps
- Minimize initial structure for each window for *N* steps
- No minimization

FEP Workflow

In this section you specify the number of windows (steps) in the perturbation, and the number of configurations in the various sampling stages.

- Split perturbation into N windows text box—Divide the perturbation into the number of steps ("windows") specified in the text box. Each window is run as a separate subjob, and can be run on a separate processor.
- Extra density at lambda termini option menu—Specify the number of points use for a higher density of lambda points at the end points of the perturbation (lambda=0 and lambda=1). The total number of windows remains the same, so the lambda values in the middle section of the perturbation are more spread out, while the lambda values at the ends are closer together. The algorithm for determining the number of points and the lambda step is as follows:
  1. The number of extra density points at each end of the range (twice the value n selected from the option menu) is subtracted from the total number of points. For example if there are 14 point total and 2 extra density points, the result is 10 points.
  2. The main lambda step is determined by dividing the remaining points evenly. In the example above, the step would be 0.1.
  3. At each end, n+1 windows are then combined and redivided into 2n+1 windows, so that there are n extra points for these steps. For the example above, the 3 windows at the beginning and the end of the range are redivided into 5 windows, each with a lambda step of 0.06.
- Solvent-only equilibration for N configurations option and text box—Select this option to perform equilibration of the solvent with the solute fixed at the given temperature and pressure, and enter the number of configurations (in thousands) for the equilibration in the text box.

- System equilibration for N configurations option and text box—Select this option to equilibrate the entire system, and specify the number of configurations (in thousands) for the equilibration in the text box. This step is performed after any solvent equilibration. All degrees of freedom are sampled.
- System averaging for N configurations option and text box—Select this option to perform sampling of the system to obtain the final results from the usual averaging techniques, and specify the desired number of configurations (in thousands) in the text box.

When you have finished making the settings, click Start to make job settings and start the job, or click Write to write the input file.

7.3 Output Files

An FEP job generates a set of output files for each λ point, *filename*`.out`, *filename*`.log`, *filename*`.mae`, and *filename*`.zmat`, where *filename* is constructed from *jobname* with various suffixes that identify the calculation stage. For each λ point, a set of output files from each MC block is written, with properties as described in Section 6.2 on page 30. The properties that are added to the Maestro file are listed in Table 7.1. The average values are the running average for the MC blocks completed to the current point. In the same way as for MC simulations, the values in the output file for the λ point are those for the last block run.

During the course of the calculation, a `.mon` file is written for each λ point, which contains a summary of the results up to the last completed simulation block. These files are copied back periodically to the job submission host by Job Control and can be used to monitor the progress of the job. They are also used to generate an intermediate summary file, *jobname*`-sum.data`, which can be used for analysis while the job is in progress. The properties available from this file are listed in Table 7.2.

The final output of the job consists of the *jobname*`.out`, *jobname*`.log`, *jobname*`-out.mae`, *jobname*`-out.zmat`, and *jobname*`-sum.data` files.

The output Maestro file, *jobname*`-out.mae`, contains the final configuration from each λ point (that is, the middle window). It also includes average properties over each window, and the static properties of the last structure. The properties that are added to the Maestro file are listed in Table 7.1; the per-block properties are for the last block.

The `.out` file contains the text output of the job. It is essentially the same as the MCPRO output file, but does not contain the Z-matrix at the beginning, and has some additional progress information.

The file *jobname*-sum.data includes a summary of properties from the entire perturbation, which can be used in the MCPRO+ Simulation Analysis panel. The properties available from this file are listed in Table 7.2 and Table 7.3.

Table 7.1. Properties found in FEP runs. In the property names, a and b are the initial and final λ values for the step.

Property	Description
Delta G <*a*->*b*> (kcal/mol)	Per-block ΔG for a perturbation.
Delta H <*a*->*b*> (kcal/mol)	Per-block ΔH for a perturbation.
Delta S <*a*->*b*> (cal/molK)	Per-block ΔS for a perturbation.
stdev Delta G <*a*->*b*> (kcal/mol)	Per-block standard deviation of the ΔG for a perturbation.
stdev Delta H <*a*->*b*> (kcal/mol)	Per-block standard deviation of the ΔH for a perturbation.
stdev Delta S <*a*->*b*> (cal/mol K)	Per-block standard deviation of the ΔS for a perturbation.
average Delta G <*a*->*b*> (kcal/mol)	Averaged ΔG over all blocks in the perturbation.
average Delta H <*a*->*b*> (kcal/mol)	Averaged ΔH over all blocks in the perturbation
average Delta S <*a*->*b*> (cal/mol K)	Averaged ΔS over all blocks in the perturbation.

7.4 Analysis of Results

The results of FEP simulations can be analyzed graphically in the MCPRO+ Simulation Analysis panel. This panel provides the means to plot various thermodynamic and energetic quantities that are generated from free-energy perturbation simulations (including relative binding affinity simulations).

To open the MCPRO+ Simulation Analysis panel, choose Analysis from the MCPRO+ submenu of the Applications menu.

The analysis is performed using the data in the *jobname*-sum.data file. The first step is to import this file. To do so, enter the file name in the Analysis file text box, or click Browse and browse to the file. The file must have the suffix -sum.data.[1] When you import the file, an

1. For Monte Carlo simulations, you can import the .out file to display data to plot, but for FEP calculations and relative binding affinity calculations, the data is in the -sum.data file

option menu is displayed and populated with the available plot types from the contents of the file. You can plot data while a simulation is in progress by importing the monitoring `-sum.data` file: in this case the plot types do not include final averaged quantities.

The properties available for plotting are listed in Table 7.2 and Table 7.3. Table 7.2 contains the properties that are available in the intermediate `-sum.data` file; Table 7.3 contains the additional properties that are generated for the final `-sum.data` file.

Table 7.2. Properties in the intermediate -sum.data file. These are also present in the final -sum.data file unless otherwise noted.

Property	Description
Job status	Status of all subjobs for FEP. Only in the monitoring `-sum.data` file.
Free Energy	Averaged free energy. Plotted as a running (cumulative average) as a function of Λ. May be plotted over multiple configurations.
System total energy	Averaged total system energy as a function of Λ. May be plotted over multiple configurations
<Solvent-solute energy>	Averaged solvent-solute energy as a function of Λ. May be plotted over multiple configurations
<Solute-solute energy>	Averaged solute-solute energy (will be zero unless have > 1 solute) as a function of Λ. May be plotted over multiple configurations.
<Bond stretch energy>	Averaged bond stretching energy of solutes as a function of Λ. May be plotted over multiple configurations.
<Angle bend energy>	Averaged angle bending energy of solutes as a function of Λ. May be plotted over multiple configurations.
<Torsional energy>	Averaged torsional energy for the solutes as a function of Λ. May be plotted over multiple configurations.
<Non-bonded energy>	Averaged non-bonded energy for the solutes as a function of Λ. May be plotted over multiple configurations. Replaced by Solute-solute non-bonded energy in the final `-sum.data` file.
<Solute-solute vdW energy>	Averaged solute-solute van der Waals energy as a function of Λ. May be plotted over multiple configurations.
<Solute-solute Coulomb energy>	Averaged solute-solute Coulomb energy as a function of Λ. May be plotted over multiple configurations.
<Solute*N*-solvent vdW energy>	Averaged solute van der Waals interaction energy with solvent as a function of Λ, for the specified solute (*N*=1, 2).
<Solute*N*-solvent Coulomb energy>	Averaged solute Coulomb interaction energy with solvent as a function of Λ, for the specified solute (*N*=1, 2).

Table 7.3. Additional properties in the final -sum.data file.

Property	Description
Enthalpy	Averaged enthalpy. Plotted as a running (cumulative) average as a function of Λ. May be plotted over multiple configurations.
Entropy	Averaged entropy. Plotted as a running (cumulative) average as a function of Λ. May be plotted over multiple configurations.
<Constraint energy>	Averaged total energy from constraints (harmonic bonds). Plotted as a function of Λ. May be plotted over multiple configurations.
<Solute-solute nonbonded energy>	Averaged solute-solute non-bonded energy as a function of Λ. Replaces <Non-bonded energy> in the intermediate `-sum.data` file.
<Solute*N*-solvent non-bonded energy>	Averaged solute non-bonded interaction energy with solvent as a function of Λ, for the specified solute (*N*=1, 2).
Solvent-solvent energy distribution	Interaction energy distribution between each solvent molecule and other solvent molecules (i.e. interaction energy versus mole fraction of solvent molecules). May be plotted over multiple configurations and Λ points.
Solute-solvent energy distribution	Interaction energy distribution between solutes and solvent molecules (interaction energy versus mole fraction of solvent molecules). May be plotted over multiple configurations and Λ points.
Solvent-solvent energy pair distribution	Energy pair distribution function for the solvent. The plot shows the average number of molecules interacting with a solvent molecule with a given energy. May be plotted over multiple configurations and Λ points.
Solute-solvent energy pair distribution	Energy pair distribution function for solute-solvent interactions. The plot shows the average number of solvent molecules interacting with the solute with a given energy. Good for revealing changes in hydrogen-bonding character over an FEP. May be plotted over multiple configurations and Λ points.

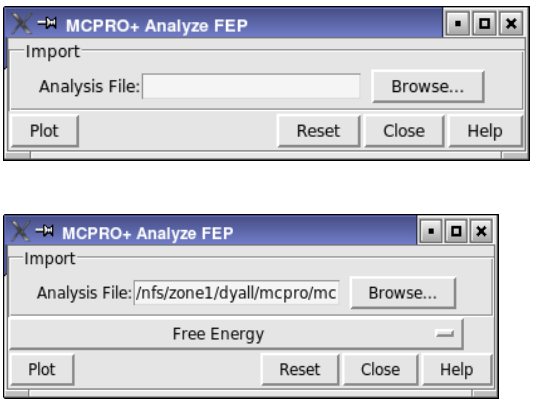


Figure 7.2. The MCPRO+ Simulation Analysis panel, before and after importing data.

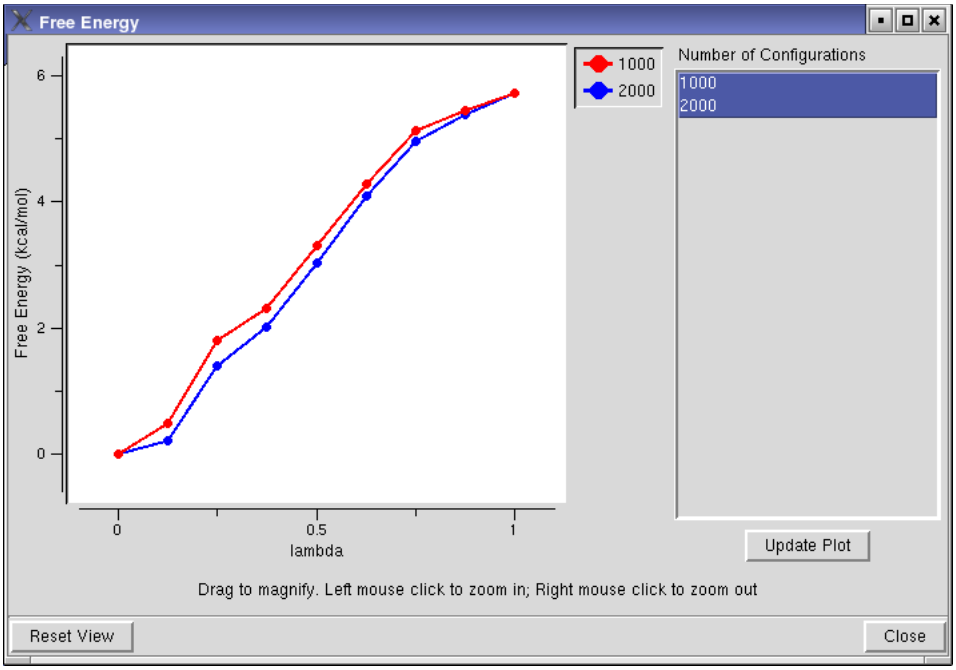


Figure 7.3. Sample Free Energy plot.

To display a plot, choose the plot type from the option menu, and click Plot. A panel opens that displays the plot for the chosen plot type. Choosing another plot type and clicking Plot opens a second panel, so you can view more than one plot at a time. The panels are labeled with the menu item for identification.

The analysis plot panels have the plotting area on the left, the plot legend in the center, and a list of data sets for plotting on the right. You can select multiple data sets in the list with shift-click and control-click. When you have selected the desired data sets, click Update Plot to redraw that plotting area with the selected data sets displayed.

You can zoom in and out on the plot with the left and right mouse buttons, or drag over an area to expand that area to the size of the plotting area. To return to the original scale, click Reset View.

If you want to clear the data from the analysis panel, click Reset. The data is cleared from the panel, including the option menu, and any plot panels that were using the data are closed.

Calculating Relative Binding Affinities

In MCPRO+ relative binding affinity calculations, the free-energy difference for the binding of two ligands to a receptor is obtained by free-energy perturbation, and the results for each ligand are subtracted to obtain a relative binding affinity. The background for this process is described in Section 2.3 on page 7. One of the advantages of the MCPRO+ procedure is that the same truncated protein is used for both ligands, which reduces the error when the difference is taken. The structural requirements for the ligands are the same as for an FEP calculation on the ligands only—see Section 7.1 on page 35 for details.

For a given pair of ligands A and B, the process involves two FEP calculations, one for the system in which ligand A is perturbed to ligand B while they are bound to the receptor (the "bound" calculation) and one for the system in which ligand A is perturbed to ligand B in the absence of the receptor (the "free" calculation).

8.1 Setting Up a Relative Binding Affinity Calculation

The setup for a relative binding affinity calculation is similar to that for a free-energy difference calculation. After setting up the model system and selecting solvation options, the parameters for the Monte Carlo simulations are set up in the Relative Binding Affinity by FEP tab. This tab has three sections:

Monte Carlo settings section

In this section you set the pressure and temperature for the ensemble, and specify the block size. Each block is run as a separate MCPRO+ calculation with the given number of configurations in the block, but the blocks are run sequentially to maintain the Markov chain.

Pre-FEP Minimization section

In this section you can request preminimization of the two structures for a specified number of steps, prior to starting the FEP workflow or at the beginning of each step. The preminimization is performed on the entire solute, and is intended to relieve any strain resulting from model system creation. You can choose one of the following options, and enter the number of steps in the appropriate text box:

- Minimize initial structure for *N* steps
- Minimize initial structure for each window for *N* steps
- No minimization

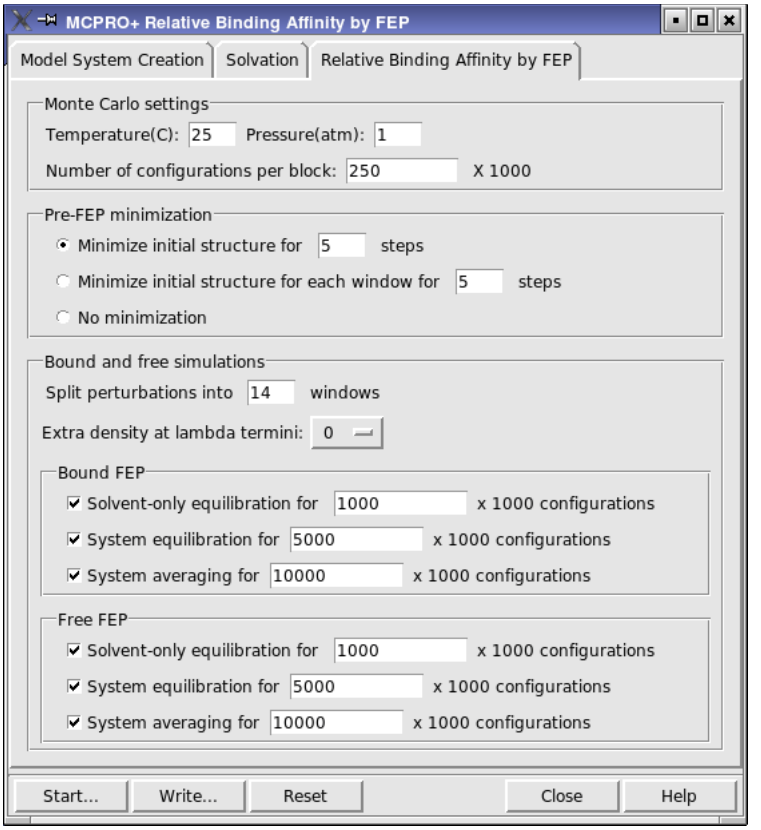


Figure 8.1. The Relative Binding Affinities tab.

Bound and Free Simulations section

In this section you specify the parameters for the sampling and the number of steps in the perturbation. The controls for the number and distribution of steps are common to both bound and free simulations. The parameters for the bound simulation and the free simulationcan be specified independently, in the Bound FEP and Free FEP sections, but the controls are identical and are described only once.

- Split perturbation into N windows text box—Divide the perturbation into the number of steps (“windows”) specified in the text box. Each window is run as a separate subjob, and can be run on a separate processor.
- Extra density at lambda termini option menu—Specify the number of points use for a higher density of lambda points at the end points of the perturbation (lambda=0 and lambda=1). The total number of windows remains the same, so the lambda values in the

middle section of the perturbation are more spread out, while the lambda values at the ends are closer together. The algorithm for determining the number of points and the lambda step is as follows:

1. The number of extra density points at each end of the range (twice the value n selected from the option menu) is subtracted from the total number of points. For example if there are 14 point total and 2 extra density points, the result is 10 points.
2. The main lambda step is determined by dividing the remaining points evenly. In the example above, the step would be 0.1.
3. At each end, n+1 windows are then combined and redivided into 2n+1 windows, so that there are n extra points for these steps. For the example above, the 3 windows at the beginning and the end of the range are redivided into 5 windows, each with a lambda step of 0.06.

- Solvent-only equilibration for *N* x 1000 configurations option and text box—Select this option to perform equilibration of the solvent at the given temperature and pressure, and enter the number of configurations (in thousands) for the equilibration in the text box.
- System equilibration for *N* x 1000 configurations option and text box—Select this option to equilibrate the entire system, and specify the number of configurations (in thousands) for the equilibration in the text box. This step is performed after any solvent equilibration. All degrees of freedom are sampled.
- System averaging for *N* x 1000 configurations option and text box—Select this option to perform sampling of the system to obtain the final results from the usual averaging techniques, and specify the desired number of configurations (in thousands) in the text box.

When you have finished making the settings, click Start to make job settings and start the job, or click Write to write the input file.

8.2 Output Files

The output files are the same as for FEP, but one set is written for the bound calculation and one set is written for the free calculation. You can analyze results for either calculation in the MCPRO+ Simulation Analysis panel. See Section 7.4 on page 43 for details.

Linear Response Calculations

The purpose of MCPRO+ linear response calculations is to provide data for the generation of linear response models for the fitting of experimental binding affinities.

The input required for a linear response calculation is a ligand file and a prepared protein model system in the same frame of reference. The structural input is the same as required for Liaison and Prime MM-GBSA calculations, but the protein must be prepared as a model system beforehand, for example by setting up the model system for an MC Sampling job for a protein only. There is no need to run the MC Sampling job or write the input files to obtain the model system, which is written at the end of the model system creation automatically.

When the job is run, MCPRO+ creates model systems for each of the ligands and their complexes. The bound Monte Carlo sampling run is similar to that done for other job types. The free ligand Monte Carlo sampling includes a step in which the ligand bond lengths and angles are held fixed and only the torsions and relative orientation are varied, with a much higher temperature used in the Metropolis criterion for ligand moves. This step and the subsequent quenching in the next equilibration allows higher-energy configurations to be sampled and thus a greater probability of moving over torsional barriers.

9.1 Setting Up the Calculations

Once the model system has been created and solvation options selected, options for the sampling can be set up in the Linear Response tab. This tab is divided into three sections:

Monte Carlo settings

In this section you set the pressure and temperature for the ensemble, and specify the block size. Each block is run as a separate MCPRO+ calculation with the given number of configurations in the block, but the blocks are run sequentially to maintain the Markov chain.

Pre-LRM Minimization

In this section you can request preminimization of the structures for a specified number of steps. The preminimization is performed on the entire solute, and is intended to relieve any strain resulting from model system creation.

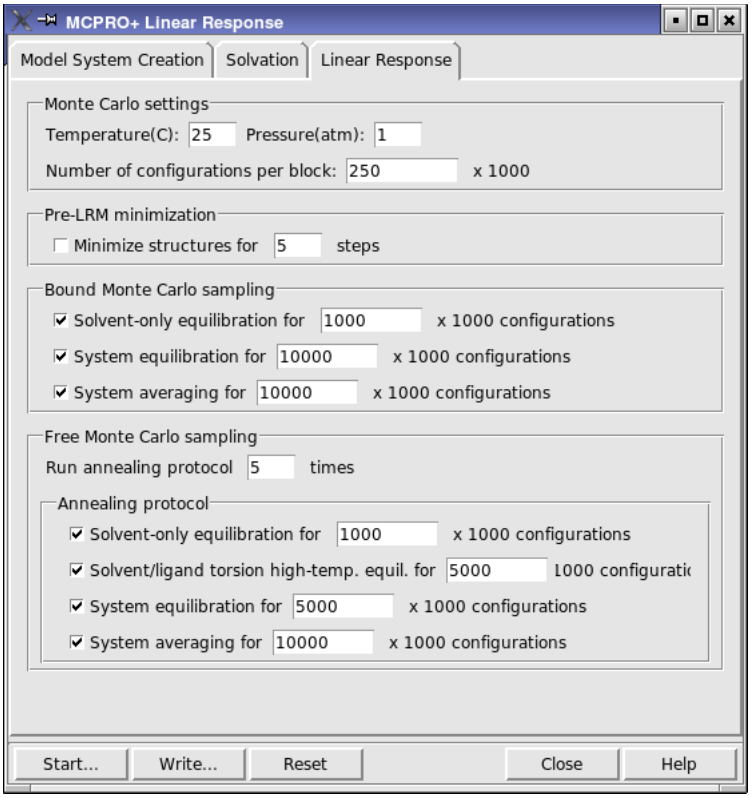


Figure 9.1. The Linear Response tab.

Bound Monte Carlo sampling

In this section you specify the number of configurations used for the sampling of the system in which the ligand is bound to the receptor. The calculations are performed in the order listed: solvent-only equilibration, system equilibration, then system averaging. You can choose to omit any of these calculations by deselecting the corresponding option. All degrees of freedom are sampled in the system equilibration

Free Monte Carlo sampling

In this section you specify the parameters for the sampling of the free ligand in solvent. This is done with an annealing protocol, which can be run multiple times, in order to obtain good Coulomb energies for the ligand-solvent system. You can specify the number of times to run the protocol.

The annealing protocol has four steps: solvent-only equilibration, solvent and ligand torsion high-temperature equilibration, system equilibration, and system averaging. For each step, you can specify the number of configurations, in thousands.

The equilibration of the solvent and the ligand torsional angles at high temperatures ensures that the ligand can sample higher-energy configurations and therefore a wider range of torsional angles. The higher temperature is quenched by the system equilibration, in which all degrees of freedom are sampled.

9.2 Output

The principal output from a linear response calculation is a Maestro file that contains the protein-ligand complexes along with the calculated properties. The results of the calculations can be incorporated into Maestro automatically. You can then use Strike to generate correlations with experimental binding affinity data, converted to free energies in kcal/mol. (See the *Strike User Manual* for information on fitting the data.)

The numerical output from a linear response calculation is also written to the file *jobname*`.csv`, which is compiled from data for the bound and the free systems. The results in this file can be used to generate a model of binding affinities. The properties that are generated and included in both the Maestro and the CSV file are listed in Table 9.1.

Table 9.1. Properties generated by linear response calculations.

Property	Description
Ligand_Number	Ligand number. Ligands are numbered by their position in the input file.
EXX-LJ	Average solute-solute van der Waals interaction energy (kcal/mol)
EXX-C	Average solute-solute Coulomb interaction energy (kcal/mol)
ESX-LJ	Average solute-solvent van der Waals interaction energy (kcal/mol) from bound simulations
ESX-C	Average solute-solvent Coulomb interaction energy (kcal/mol) from bound simulations
dElj	Difference in van der Waals solute-solute and solute-solvent interaction energies between bound and free simulations. Elj=bound EXX-LJ+(bound ESX-LJ-free ESX-LJ) (kcal/mol)
dEc	Difference in average Coulomb solute-solute and solute-solvent interaction energies (kcal/mol) between bound and free simulations. Ec= bound EXX-C+(bound ESX-C – free ESX-C)

Table 9.1. Properties generated by linear response calculations. (Continued)

Property	Description
dSASA	Difference in average solvent accessible surface area ($Å^2$) of ligand between bound and free simulations. dSASA = bound SASA - free SASA.
dFOSA	Difference in average hydrophobic solvent accessible surface area ($Å^2$) of ligand between bound and free simulations.
dFISA	Difference in average hydrophilic solvent accessible surface area ($Å^2$) of ligand between bound and free simulations.
dARSA	Difference in average aromatic solvent accessible surface area ($Å^2$) of ligand between bound and free simulations.
dWPSA	Difference in average weakly polar solvent accessible surface area ($Å^2$) of ligand between bound and free simulations.
#RB	Number of ligand rotatable bonds.
dHBdon_solvent	Difference in the average number of ligand hydrogen-bond donor sites making hydrogen bonds to solvent between bound and free simulations.
dHBdon_solute	Difference in the average number of ligand hydrogen-bond donor sites making hydrogen bonds to solute between bound and free simulations
dHBdon	Difference in the average total number of hydrogen bonds made by ligand hydrogen-bond donor sites between bound and free simulations.
dHBacc_solvent	Difference in the average number of ligand hydrogen-bond acceptor sites making hydrogen bonds to solvent between bound and free simulations.
dHBacc_solute	Difference in the average number of ligand hydrogen-bond acceptor sites making hydrogen bonds to solute between bound and free simulations
dHBacc	Difference in the average total number of hydrogen bonds made by ligand hydrogen-bond acceptor sites between bound and free simulations
dHBtot	Difference in the average total number of hydrogen bonds made by the ligand between bound and free simulations
dEint	Difference in the average internal energy (kcal/mol) of the ligand between bound and free simulations
#SatAmine	Number of saturated amine groups in the ligands
#Nitro	Number of nitro groups in the ligands
#Acid	Number of acid groups in the ligands

Examining Structure-Activity Relationships

Structure-activity relationships can be obtained with MCPRO+ by performing free-energy perturbation calculations for a series of related ligands. While it would be possible to use the Relative Binding Affinity by FEP panel and run separate calculations for each pair of ligands, MCPRO+ provides interfaces in which you can create a series of ligands from a predetermined set of fragments, and run all the calculations together. These interfaces are the Ligand Functional Group Mutation by FEP panel and the Ring Atom Mutation by FEP panel, which you open from the MCPRO+ submenu of the Applications menu.

10.1 Ligand Functional Group Mutation

For ligand mutation, the ligand series is generated from an initial ligand structure by replacing a group of your choice (the *substitution group*) with each fragment in a set of fragments that you choose from a predetermined list. The fragments are already set up as required for an MCPRO+ free-energy perturbation calculation.The original ligand is not included in the FEP calculations: it merely serves to provide a scaffold to which the fragments can be attached.

Once the ligands are defined, FEP calculations are set up to determine the binding affinity relative to a reference ligand, which you can choose from among the generated ligands. If you do not choose a reference ligand, it is chosen by minimizing the number of FEP calculations required to calcuate binding affinities for all the ligands. For each ligand, several FEP calculations may be required to transform the reference ligand into the current ligand. For example, transforming methyl to propyl is done in two steps, from methyl to ethyl, then from ethyl to propyl.

The first step is to choose the protein-ligand complex that will form the basis of the series of ligands, and display it in the Workspace. The complex need not be a single project entry, but can consist of several entries if need be. It is advisable to use a ligand that is well-positioned with respect to the protein, such as in a native complex, or a ligand docked with Glide.

Once you have displayed the complex and opened the Ligand Functional Group Mutation by FEP panel, the next step is to identify the attachment bond, which is the bond between the ligand core and the substitution group. When the panel opens, the Pick the attachment bond option is active, and you can pick a bond in the ligand structure in the Workspace to define the attachment bond. You should click the bond on the end nearest to the substitution group. A green arrow is placed over the bond, pointing to the substitution group. If you accidentally click on the wrong end of the bond, you can click again until the bond is correctly identified.

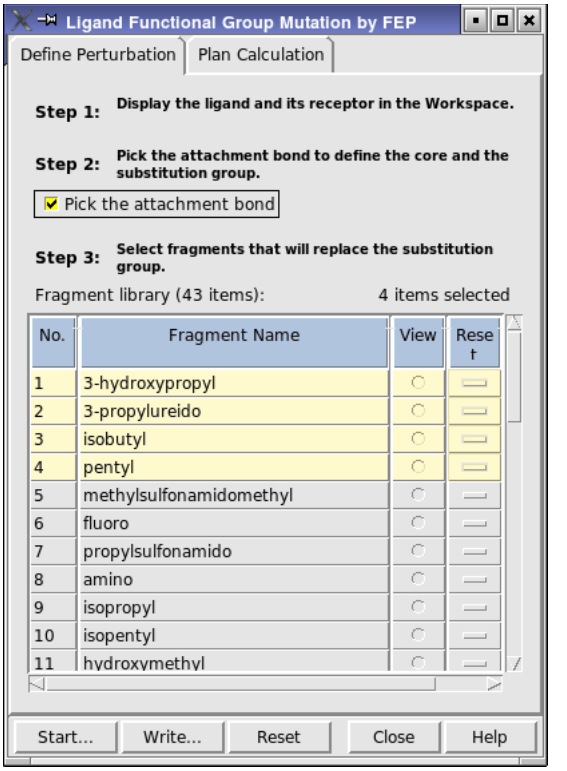


Figure 10.1. The Define Perturbation tab of the of the Ligand Functional Group Mutation panel.

When the attachment bond is defined, you can select the fragments that you want to use to generate the ligand series. You do this by selecting rows in the table in the lower part of the panel. Shift-click and control-click work in the usual way for selection of table rows.

To view a ligand, click in the View column for that ligand. When the ligand is displayed, the fragment is added in a standard orientation. This orientation might have clashes with the protein. You can adjust the fragment (and indeed any part of the ligand) using the adjustment tools, which are available from the Adjust button menu on the toolbar.

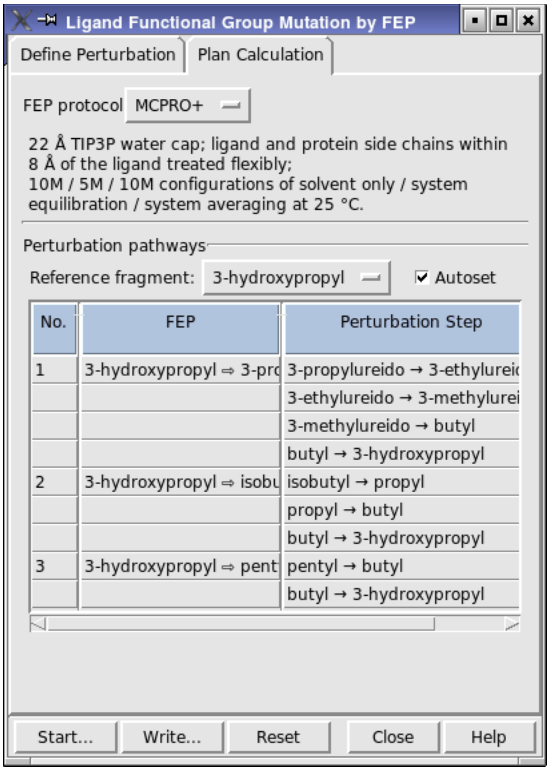


Figure 10.2. The Plan Calculation tab of the Ligand Functional Group Mutation by FEP panel.

You might want to display contacts between the ligand and the receptor while you are adjusting the structure. To do this, open the Measurements panel from the Tools menu, and in the Contacts tab, pick the ligand for Atom Set 1 and the receptor for Atom Set 2. The adjustments you make to the ligand are stored and used in the FEP calculations. If you want to revert to the standard orientation, click the button in the Reset column.

The calculation is performed with a protocol that defines the parameters of the calculation. This protocol is described briefly in the Plan Calculation tab. The protocol can be chosen from the FEP protocol option menu. There are only two choices: MCPRO+ and User defined. If you choose User defined, a text box and Browse button are displayed, in which you can select a protocol file, which has the extension `.msj`.

If you want to select a reference ligand, choose the desired reference fragment from the Reference Fragment option menu. The relative binding affinities are then all expressed with respect to the value for this ligand. To automatically set up the reference fragment, click Autoset.

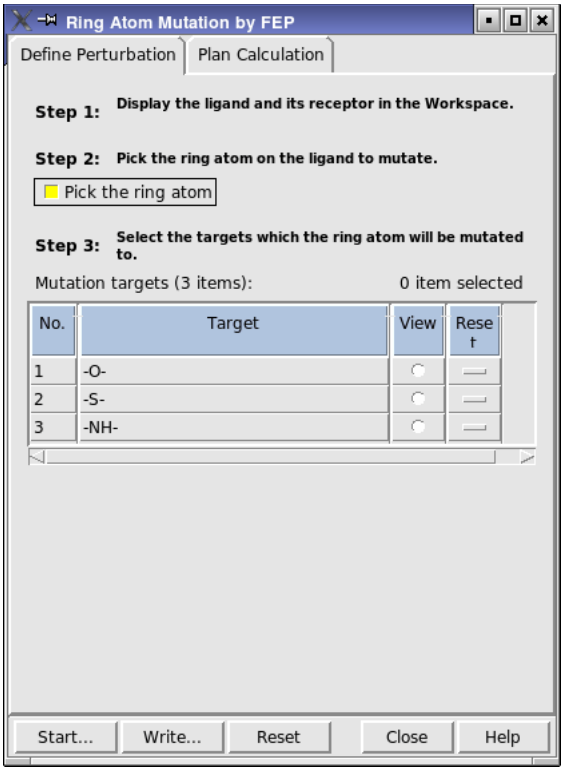


Figure 10.3. The Define Perturbation tab of the of the Ring Atom Mutation by FEP panel.

The perturbation pathways are displayed in a table in the Plan Calculations tab, which you can use to set up parameters for the FEP calculations. Each pathway takes the reference ligand to the relevant ligand via a predetermined set of intermediates. These intermediates have been chosen to minimize the size of the perturbations and hence increase the accuracy of the final result. When the actual calculations are performed, a unique, minimum set of perturbations is determined using graph theory, and it is these calculations that are run. The number of calculations is therefore independent of the choice of reference ligand.

10.2 Ring Atom Mutation

For rign atom mutation, the ligand series is generated from an initial ligand structure by replacing a ring atom of your choice that is only attached to two other ring atoms and to hydrogen with each of a set of targets that you choose from a predetermined list. The targets are already set up as required for an MCPRO$^+$ free-energy perturbation calculation.The original ligand is included in the FEP calculations.

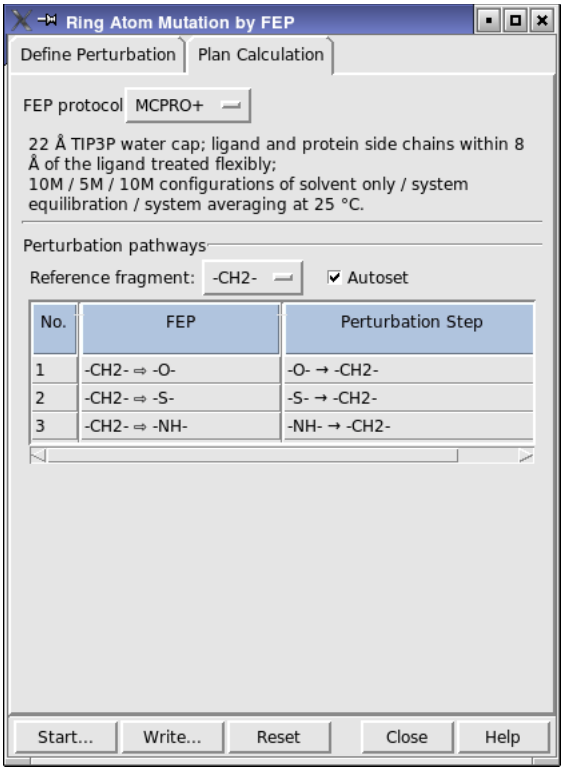


Figure 10.4. The Plan Calculation tab of the Ring Atom Mutation by FEP panel.

The first step is to choose the protein-ligand complex that will form the basis of the series of ligands, and display it in the Workspace. As for ligand functional group mutation, the complex need not be a single project entry, but can consist of several entries if need be. It is advisable to use a ligand that is well-positioned with respect to the protein, such as in a native complex, or a ligand docked with Glide.

Once you have displayed the complex and opened the Ring Atom Mutation by FEP panel, the next step is to identify the ring atom. When the panel opens, the Pick the ring atom option is active, and you can pick a ring atom in the ligand structure in the Workspace. The atom is rendered in ball-and-stick representation. A green arrow is placed over the bond, pointing to the substitution group. If you accidentally pick the wrong atom, you can pick again until the correct atom is picked.

When the ring atom is defined, you can select the targets that you want to use to generate the ligand series. You do this by selecting rows in the table in the lower part of the panel. Shift-click and control-click work in the usual way for selection of table rows. To view a ligand, click in the View column for that ligand.

The calculation is performed with a protocol that defines the parameters of the calculation. This protocol is described briefly in the Plan Calculation tab. The protocol can be chosen from the FEP protocol option menu. There are only two choices: MCPRO+ and User defined. If you choose User defined, a text box and Browse button are displayed, in which you can select a protocol file, which has the extension `.msj`.

If you want to select a reference ligand, choose the desired reference fragment from the Reference Fragment option menu. The relative binding affinities are then all expressed with respect to the value for this ligand. To automatically set up the reference fragment, click Autoset.

10.3 Running the Job

Once you have made all settings, click Start. The Start dialog box opens, in which you can select a host on which to run the job. The job initiates a number of subjobs for each FEP, and each FEP step can be run on a separate procesor, so you may want to choose a multiprocessor host. When you click Start in the Start dialog box, the job is submitted to the selected host. You can monitor the progress of the job in the Monitor panel, which you can open from the Applications menu.

The initiation of the subjobs can take some time, because a model system has to be created for each ligand and each intermediate (if there are any). Each model system creation and subsequent subjob submission can take a minute or two.

When the jobs have finished, the final structures are returned to the Project Table with the relative binding affinity as a property. The results are also tabulated in the *jobname*`.log` file, which you can view in the Monitor panel. These results include the free energy differences for the full set of perturbations with the intermediates, as well as the final results for the selected fragments.

Instead of running the job immediately, you can write out the command and structure files by clicking Write. You can then run the job from the command line, with the `mcpro` command. Before you run the job, you must generate the input files with the `autoper` utility, as follows:

`$SCHRODINGER/utilities/autoper` *jobname*`.cmd` *jobname*`_pv.mae`

Running MCPRO+ from the Command Line

There may be occasions on which you want to run MCPRO+ from the command line. This chapter describes the command-line tools provided with MCPRO+ that allow you to run various jobs.

MCPRO+ jobs run a version of the MCPRO program that has been modified in a few places with respect to the Jorgensen distribution.

All jobs run under the Job Control facility, and accept the standard Job Control options and `-WAIT` and `-LOCAL`. These options are described in Section 2.3 of the *Job Control Guide*. You should use the Job Control facility to manage your jobs, including killing them when necessary.

11.1 Input Files

The input file for an MCPRO+ job is a file named *jobname*`.inp`. This file is written when you click Write in the MCPRO+ panels. In addition, you must have a model system, which consists of a *jobname*`.mae` file and a *jobname*`.zmat` file. All other input files required by the MCPRO program are generated from these files.

The input file consists of keyword-value pairs that define the type of calculation to be run. These keyword-value pairs are divided into blocks. Only one keyword-value pair is allowed per line. The keyword and the value are separated by an = sign with no spaces. Comments can be included in the file, by placing a # character at the beginning of the line.

There are two essential blocks in the MCPRO+ input file, a workflow block and one or more sampling blocks. Parameter blocks can also be included: one or more of these are written automatically by the MCPRO+ GUI per sampling block. The workflow block contains options that are carried throughout the simulation or used as the basis for setting sampling or parameter block options. Blocks are separated by a blank line. For sampling blocks the first keyword must be `blockname`. For a parameter block the first keyword must be `newparfile`. Each sampling block must set the `blockname` and `parameter` options. Parameters specified in a parameter block will be written to the file named by `newparfile`.

The format of the input file is demonstrated with the following MCPRO+ input file. This simulation specifies1M configurations of MC sampling of solvent only, followed by 5M configurations of equilibration and 10M configurations of averaging.

```
# MCPRO+ input file written by Maestro
# To perform Complex (bound) FEP calculation for ddG
jobtype=mc
zmatrix=C02_C01_c12_f8-comp.zmat
strfile=C02_C01_ddg_out.mae
outfile=C02_C01_ddg.out
maestrfile=C02_C01_c12_f8-comp.mae
configurations=250000

# Solvent-only sampling block
blockname=C02_C01_ddg_bound_solEq
parameter=C02_C01_ddg_bound_solEq0.par
numtotalconfigs=1000000
parameter2=C02_C01_ddg_bound_solEq.par

# Equilibration sampling block
blockname=C02_C01_ddg_bound_Eq
parameter=C02_C01_ddg_bound_Eq.par
numtotalconfigs=5000000

# Averaging sampling block
blockname=C02_C01_ddg_bound_Ave
numtotalconfigs=10000000
parameter=C02_C01_ddg_bound_Ave.par

# Parameters for file C02_C01_ddg_bound_solEq0.par
newparfile=C02_C01_ddg_bound_solEq0.par
wkc=150
nrot1=4
caprad=24
adels1=0.0
nvchg=999999
p=1
svmod1=TIP3P
nschg=250001
capat=CAP
t=25
soluorg=ZMAT
nschg2=250001
rdels1=0.00
izlong=1
solvorg=BOXES
nrot2=2783
ncent1=4
ncent2=2783
adels2=-1.0
rdels2=-0.15
```

Comments start with #

Workflow block

Sampling block

Sampling block

Sampling block

Parameter block

```
# Parameters for file C02_C01_ddg_bound_solEq.par
newparfile=C02_C01_ddg_bound_solEq.par
wkc=150
nrot1=4
caprad=24
adels1=0.0
nvchg=999999
p=1
svmod1=TIP3P
nschg=250001
capat=CAP
t=25
nschg2=250001
rdels1=0.00
nrot2=2783
ncent1=4
ncent2=2783
adels2=-1.0
rdels2=-0.15
```

Parameter block

```
# Parameters for file C02_C01_ddg_bound_Eq.par
newparfile=C02_C01_ddg_bound_Eq.par
wkc=150
nrot1=4
caprad=24
adels1=0.0
nvchg=999999
p=1
svmod1=TIP3P
capat=CAP
t=25
rdels1=0.00
nrot2=2783
ncent1=4
ncent2=2783
adels2=-1.0
rdels2=-0.15
```

Parameter block

```
# Parameters for file C02_C01_ddg_bound_Ave.par
newparfile=C02_C01_ddg_bound_Ave.par
wkc=150
nrot1=4
caprad=24
adels1=0.0
nvchg=999999
p=1
svmod1=TIP3P
capat=CAP
t=25
rdels1=0.00
```

Parameter block

```
nrot2=2783
ncent1=4
ncent2=2783
adels2=-1.0
rdels2=-0.15
```

Keywords for the workflow block are given in Table 11.1. The keywords are divided into two groups: workflow settings, and file settings. Keywords for the sampling blocks are given in Table 11.2, and for the parameter blocks in Table 11.3. Required keywords are noted in the tables.

Table 11.1. Keywords for the workflow block.

Keyword	Description
Workflow settings	
`jobtype`	Required. Allowed values are `mini`, `mc`, and `sp`. Set the type of job to be run: minimization, monte carlo sampling, or single point energy calculation. Required.
`comment`	User-provided comment string that is added to the command file as the comment in the header. Default is the job name.
`configurations`	For an MC run, sets the number of configurations per MCPRO block and provides the default for `numconfigspersubblock`. For a minimization run, sets the number of minimization steps to be run. Default: 200000.
`verbosity`	Set the verbosity level for all MCPRO subjobs. Allowed values: 0,1 Typical settings used 2 Also print solute coordinates 3 Includes high-energy nonbonded pairs and SASAs for all solute atoms 4 Includes residue-residue and solvent-residue lists (use sparingly) 5 Debugging purposes only Default: 1
`rc0`	L setting for reference system in FEP. No default value. Required for a single L point run along with `rc1` and `rc2`. Must not be included the input file to `mcpro_fep` or `mcpro_ddg`.
`rc1`	L setting for 1st perturbed system in FEP. No default value. Required for a single L point run along with `rc0` and `rc2`. Must not be included the input file to `mcpro_fep` or `mcpro_ddg`.
`rc2`	L setting for 2nd perturbed system in FEP. No default value. Required for a single L point run along with `rc0` and `rc1`. Must not be included the input file to `mcpro_fep` or `mcpro_ddg`.

Table 11.1. Keywords for the workflow block. (Continued)

Keyword	Description
File settings	
zmatrix	Z-matrix component of a model system for the solutes. Required. File must exist.
parameter	Parameter file with potential function parameters and variables for the simulation. This file is written by the driver if a parameter block is identified.
outfile	Output text file with all results of a simulation.
strfile	Output structure file.
infile	Restart file. For the first job in an MC sampling block it is generally written as output, then used as input for future jobs in that block.
bangpar	Bond and angle parameter file. From data directory $SCHRODINGER/mcpro-v*version*/data/.
waterbox	File with water box coordinates. From data directory $SCHRODINGER/mcpro-v*version*/data/.
org1box	File with organic solvent box coordinates. From data directory $SCHRODINGER/mcpro-v*version*/data/.
org2box	File with additional organic solvent box coordinates. From data directory $SCHRODINGER/mcpro-v*version*/data/.
upfile	Ouput file with the results for each run needed to compute the global averages.
savefile	Used to store all coordinates periodically for later analysis (period specified by saveeverynconfigs)
avefile	Averages for distribution functions that may be plotted.
sumfile	Short version of the output file.
maestrfile	Input template Maestro file used to create output structure files in Maestro format when strfmt=MAE. Only required if strfmt=MAE.

Table 11.2. Monte Carlo sampling block keywords

Keyword	Description
blockname	Required; string. Sets a short name to be used to distinguish a block of MC sampling. Typical values are eq*integer* and ave*integer*. All blocknames in an input file must be unique. A blockname starts a special group of commands used to setup a single MC sampling block.
numtotalconfigs	Required; integer. Number of configurations to be run in the block of MC sampling (i.e. in full equilibration or averaging loop).
numconfigspersubblock	Integer. Number of configurations to be run per subblock of MC sampling. Default is the value given by the configurations keyword, and typically this value does not need to be changed.
saveeverynconfigs	Integer. Number of configurations between writing to the save file. Default is max(1,(numtotalconfigs/numconfigspersubblock)/4.0)
parameter2	Name of parameter file with potential function parameters and variables for the simulation. This file is written if a parameter block is identified. The parameter2 file is used for all but the first mcpro run when found inside an MC Sampling block. Most commonly used to change the input read, i.e. ZMAT to BOXES.

Table 11.3. Parameter block keywords.

Keyword	Description
newparfile	Required. Name of parameter file that will be written for this parameter block. This keyword must be at the start of the parameter block, which must end with a blank line. Each parameter block must have a unique value of newparfile.
svmod1	Sets the primary solvent choice. Allowed values are: GBSA, TIP3P, TIP4P, CH3OH, DMSO, THF, CCL4, ARGON, CH3CN, MEOME, C3H8, MECL2, CHCL3, NONE. Only GBSA, TIP3P, TIP4P, or NONE may be used with strfmt=MAE. Default: NONE
soluorg	Specify where the initial solute geometry comes from. Allowed values are: ZMAT, PDB, MIND, IN, ZIN. Value is case-insensitive. It is typical to use ZMAT for first subjob and then ZIN for further subjobs in an MC sampling block. Default: ZIN.
solvorg	Specify where the initial solvent geometry comes from. Allowed values: BOXES, IN, NONE. Value is case-insensitive. It is typical to use BOXES for first subjob in equilibration and then IN for all further subjobs. Default: IN.
capat	When a solvent cap is requested, capat designates the name of the solute atom that defines the center of the cap. Allowed values: cap atom name or NONE. Value is case-insensitive. The cap atom is typically called CAP in the Z-matrix. Default: NONE.

Table 11.3. Parameter block keywords.

Keyword	Description
caprad	Real. Water molecules will be added up to caprad Å from capat. Default: 0.0
capfk	Sets the restoring force constant in kcal/mol Å^2 for the half-harmonic restraining potential required to keep the water cap in place. Default:1.5.
optimizer	Selects the optimizer to be used for minimization. Allowed values are: SIMPLX, FLEPOW, CONJUG. Value is case-insensitive. Default: FLEPOW.
tolerance	Energy tolerance in kcal/mol for convergence of the optimization. Default: 0.0001.
newzmat	If defined the final Z-matrix from an run is appended to the sum file. Allowed values: NEWZMAT or blank.
nmol	Specifies the number of solvent molecules to be included in the solvent box. The ibox – nmol highest-energy solvent molecules are removed. If set to 9999, a solvent box is automatically created (see bcut). Default: 0.
ibox	Number of molecules in the desired stored solvent box. For water the allowed values are 216, 250, 267, 324, 400, 512, and 750. (See the MCPRO manual for other solvents) This keyword is used to select the proper solvent box. Default: 0.
bcut	Box dimension for automatic solvent box creation. If nmol is set to 9999, a solvent box is automatically created. The box extends bcut Å beyond the farthest solute atom in the positive and negative *x*, *y*, and *z* directions. If you use this option, you must set irecent=1, izlong=1, and use a value of nvchg greater than 300,000. Default: 7.0.
ncent1	Atom number of the atom used to define the center of solute 1 for centering, translation, and preferential sampling. Solute 1 is the protein in a protein-ligand complex. No default.
ncent2	Atom number of the atom used to define the center of solute 2 for centering, translation, and preferential sampling. Solute 2 is the ligand in a protein-ligand complex. No default.
nrot1	Atom number of the atom that solute 1 is rotated about in rigid-body rotation. Typically the same as ncent1. Default: value of ncent1.
nrot2	Atom number of the atom solute 2 is rotated about in rigid-body rotation. Typically the same as ncent2. Default: value of ncent2.
noback	Set to 1 if a protein backbone is to be held fixed; set to 0 if the backbone is allowed to move. Allowed values are 0 and 1. Used to set rdels1 and adels1 to zero to prevent rigid-body rotation and translation of the protein. Default: 1.
rsolv	Probe radius for SASA calculation in angstroms. Default: 1.4
norrup	Controls variation of nonbonded residue pair list. Set to 0 if the residue pair list is allowed to change over the course of a simulation. Set to 1 if the initial residue pair list is to be used throughout the simulation. Allowed values: 0,1. Default: 1.

Table 11.3. Parameter block keywords.

Keyword	Description
ifixrr	Determines when the intermolecular residue-residue list is computed. Allowed values: 0, 1. If set to 1, the list is only computed at initialization of a sampling block; in this case norrup is set to 1. This value helps to ensure that a ligand sees the same protein environment for an entire simulation. Allowed values: 0,1. Default: 1.
irecent	If set to 1, recenters solutes in the middle of the simulation box. Purely cosmetic. Allowed values: 0,1. Default: 1.
indsol	If set to 1 let solutes move independently. Default = 1.
izlong	If 1 solute(s) are oriented with the longest axis along the z-direction. Only do in first subjob of a sampling block. Allowed values: 0,1. Default: 1.
maxovl	If set to 1 in an FEP simulation have the perturbed solutes maximally overlaid on the reference solute. This helps reduce the noise. For proteins maxovl=0 to avoid global movement. Allowed values: 0,1. Default: 0.
noxx	For use with multiple solutes. If noxx=1 solute-solute interactions are not evaluated. Allowed values: 0,1. Default: 0.
noss	For use with solvent copies. If noss=1 solvent-solvent interactions are not evaluated. Allowed values: 0,1. Default: 0.
nobndv	If set to 1 variable bonds are not varied. Allowed values: 0,1. Default: 0.
noangv	If set to 1 variable angles are not varied. Allowed values: 0,1. Default: 0.
noebd	If set to 1 only bonds declared as variable and additional in the z-matrix are considered to be covalent bonds for determining covalent neighbors. Allowed values: 0,1. Default: 1.
nrdf	Maximum number of solvent-solvent radial distribution functions to be computed. Allowed values: 0,1, 2, 3. Default: 0.
ibdpdb	Set to 1 to avoid falsely designating two atoms in different residues as being bonded due to poor non-bonded distance. Allowed values: 0,1. Default: 0.
iddd	Set to 1 to use a distance-dependent dielectric in gas-phase MC simulations or CG minimization. ε is then iddd*r_{ij}. Allowed values: 0,1. Default: 0.
stren	Ionic strength in mol/dm^3 for GB/SA calculations. Default: 0.0.
rdlmin	Minimum value for RDFs in angstroms. Default: 1.350.
rdlinc	Increment unit for RDFs in angstroms. Default: 0.10.
eprmin	Minimum value for solvent-solvent energy pair distribution in kcal/mol. Default: −10.25.

Table 11.3. Parameter block keywords.

Keyword	Description
eprinc	Increment unit for solvent-solvent energy pair distribution in kcal/mol. Default: 0.50.
edmin	Minimum value for solvent-solvent energy distribution in kcal/mol. Default: –40.50.
edinc	Increment unit for solvent-solvent energy distribution in kcal/mol. Default: 1.0.
essmin	Minimum value for solvent-solute energy distribution in kcal/mol. Default: –15.2.
essinc	Increment unit for solvent-solute energy distribution in kcal/mol. Default: 0.40.
ebsmin	Minimum value for solute-solvent energy distribution in kcal/mol. Default: –151.0
ebsinc	Increment unit for solute-solvent energy distribution in kcal/mol. Default: 2.00.
nvchg	Every nvchg configurations an attempt is made to adjust the system's volume. If set to 99999 an NVT simulation is run. If set to 0 the value is determined automatically. Default: 0.
nschg	Every nschg configurations an attempt is made to move any solute. If set to 0 the value is determined automatically. Default: 0.
nschg2	Every nschg2 configurations an attempt is made to move the second solute. May be used to bias sampling. If set to 0 the value is determined automatically. Default: 0.
nconrot	Every nconrot configurations a concerted motion move for the biomolecule's backbone is attempted. The Z-matrix must have been built in pepz including the set conrot command. Default: 0.
nconsv	Write coordinates to savefile every nconsv configurations. Default: 999999.
nbuse	If set to 1 use solvent-solvent neighbor lists. Allowed values: 0,1. Default: 1.
maxvar	Maximum number of variable bonds, angles, and dihedrals that will be varied on an attempted solute move. If maxvar is set to 0, a value of 15 is used. Default: 4.
nsafrq	Frequency for reevaluating the SA term for GB/SA solvation in MC. If nsafrq is set to 0, a value of 30 is used. Default: 30.
vdel	Sets the size of volume moves attempted. Default: 0.00.
wkc	Constant in preferential solvent sampling. Default: 0.00.
rdel	Ranges for attempted translation in Å of solvent molecules. Assigned automatically for a value of 0.00. Default: 0.00.
adel	Ranges for attempted rigid-rotation in degrees of solvent molecules. Assigned automatically for a value of 0.00. Default: 0.00.

Table 11.3. Parameter block keywords.

Keyword	Description
rdels1	Ranges for attempted translation of solute 1 in Å. Default: −1.00 which will result in automatic assignment of this term.
adels1	Ranges for attempted rigid-body rotation of solute 1 in degrees. Default: −1.00 which will result in automatic assignment of this term.
rdels2	Ranges for attempted translation of solute 2 in Å. Default: −1.00 which will result in automatic assignment of this term.
adels2	Ranges for attempted rigid-body rotation of solute 2 in degrees. Default: -1.00 which will result in automatic assignment of this term.
rcut	Cutoff distance in Å for solvent-solvent interactions. Default = 9.0
scut	Cutoff distance in Å for solute-solvent interactions. Default = 9.0
cutnb	Cutoff distance in Å for intrasolute interactions. Default = 100.0
t	Temperature in Celsius. Default: 25.
p	Pressure in atmospheres. Default: 1.00.
locheat	Local heating in Celsius. Default is the temperature.
diel	Dielectric constant, ε, for solute optimization and continuum simulations. Default: 1.0.
scl14c	Coulombic scaling factors for 1,4-intramolecular non-bonded interactions. If both scl14c and scl14l are greater than 99999, then 1,4 interactions are ignored. Default: 2.0.
scl14l	Lennard-Jones scaling factors for 1,4-intramolecular non-bonded interactions. If both scl14c and scl14l are greater than 99999 then 1,4 interactions are ignored. Default: 2.0.
strfmt	Format of the output structure file. Allowed values are: PDB, PDB2, PDBB, MAE, MIND. strfmt=MAE requires the maestrfile keyword to be set. Default: MAE.
isolec	Solute-solvent and solute-solute energies are decomposed into coulomb and LJ components for solute isolec. Used to identify the ligand; required for linear response runs to generate the LRM terms. Allowed values: 1, 2. Default: 2.
icalc	Sets icalc. Allowed values: 0,1,9. Otherwise set by location in simulation block.
newrun	Sets newrun. Allowed values: 0,1. Otherwise set by location in simulation block

11.2 Output Files

For basic MC sampling runs, the output of an MCPRO+ run is the series of `.out` files whose names are specified by setting either the `outfile` keyword in the workflow block or the `blockname` keyword in a sampling block. Typically there will be (total number of configurations run)/(number of configurations per block) `.out` files. The `.out` file contains energetic information from the current sampling block and averaged results. Information on the properties in these files (and the output Maestro files) is given in Chapter 5 for minimization and Section 6.2 on page 30 for MC simulations.

For FEP and DDG (relative binding free energy) runs, important data from each FEP is compiled in a file named *jobname*`-sum.data`. This file contains tables of interesting properties and the data is used in the MCPRO+ Simulation Analysis panel to view plots of the many properties as a function of L and number of configurations. Further information is given in Section 7.3 on page 42.

For linear response simulations a *jobname*`.csv` file is generated for each ligand. These CSV files are then compiled together into a CSV file with descriptors from all ligands. The linear response terms are defined in Table 9.1 on page 53.

11.3 Minimizations and Basic Monte Carlo Simulations

Minimizations and basic Monte Carlo simulations can be run with the `mcpro` command:

`$SCHRODINGER/mcpro [`*options*`]` *input-file*

Apart from the Job Control options, there are only two options: `-h` or `-help` to display the usage message, and `-v`, to display the MCPRO version.

There are three forms for the input file specification:

- *jobname*`.inp`—Use the MCPRO+ input file. When this form is used, the C shell script and other MCPRO input files are generated and used to run the job.
- *jobname*`.csh`—Use an existing C shell script. The other MCPRO input files must already exist.
- MCPRO arguments—Use the arguments to the MCPRO program as specified in the *MCPRO Version 2.0 User's Manual*. All the input files must already exist.

Each MCPRO+ job produces a *jobname*`.log` file, which contains the execution log, and a *jobname*`.out` file, which is the MCPRO text output file, without the initial Z-matrix and with status records added. The output structures are written to *jobname*`-out.mae` and *jobname*`-out.zmat`. Structure output files are written at the end of each block in the simulation.

Table 11.4. MCPRO+ output files

File suffix	Description
`.log`	Execution log.
`.out`	MCPRO output file.
`-out.mae`	Maestro structure file. Includes explicit waters.
`-out.zmat`	Z-matrix file.

11.4 Free-Energy Perturbation Simulations

Free energy perturbation (FEP) calculations can be run with `mcpro_fep`. FEP jobs run a set of subjobs, one for each λ value, either serially or simultaneously. Each of these subjobs runs the `mcpro` command, and generates its own set of output files. These output files are labeled with the subjob type and number. The output structure file for the job contains the starting and ending structure for each of the windows. The FEP job itself does not have a text output file, only an execution log file. This log file does contain results.

A summary of the results is written at the end of each block and is copied to the job submission host, where it can be used to monitor the progress of the simulations. This file is named *jobname*`.mon`. At the end of the simulation, a summary is written to the file *jobname*`.sum` with the final results. A summary file of FEP results, over all λ values, is written periodically to *jobname*`-sum.data`. This file can be read in the MCPRO+ Simulation Analysis panel and used to display plots of the data.

Syntax:

```
$SCHRODINGER/mcpro_fep [options]
```

The options are described in Table 11.5.

Table 11.5. Options for the mcpro_fep command.

Option	Description
`-b` *verbose*| `--verbose=`*verbose*	Verbosity level (0-3 with default of 0)
`-c`|`--compile`	Compile results from completed a completed FEP run only
`-h`|`--help`	Display usage message and exit
`-i` *infile*| `--infile=`*infile*	MCPRO+ input file
`-m`|`--mini`	Minimize initial structure prior to running FEP

Table 11.5. Options for the mcpro_fep command. (Continued)

Option	Description
`-n` *npert*|`--npert=`*npert*	Total number of perturbations to generate complete FEP
`-p` *pert*|`--pert=`*pert*	Single perturbation to be expanded
`-t`|`--notrap`	Return complete error messages
`-v`|`--version`	Print version of the `mcpro_fep` program
`-w`|`--write`	Only write the specified input files

11.5 Relative Binding Affinity Simulations

The bound and free free-energy perturbation simulations required to obtain a relative free energy difference can be run with `mcpro_ddg`. From an input file for the free simulation and an input file for the bound simulation, this program creates the necessary input files to run multiple windows at different λ points, and runs the simulations serially or simultaneously. Output is produced for both bound and free FEPs, including two *name*`-sum.data` files.

Syntax:

`$SCHRODINGER/mcpro_ddg [`*options*`] -j` *jobname* `-n` *bound-perts* `-b` *bound-file* `-m` *free-perts* `-f` *free-file*

The options are described in Table 11.6.

Table 11.6. Options for the mcpro_ddg command.

Option	Description
`-b` *infile*| `--bound=`*infile*	MCPRO+ input file for the first (bound) perturbation. Required.
`-c`|`--compile`	Compile results from a completed DDG run only
`-f` *infile2*| `--free=`*infile2*	MCPRO+ input file for the second (free) perturbation. Required.
`-h`|`--help`	Display usage message and exit
`-i`|`--mini`	Minimize initial structures prior to running FEPs
`-j` *jobname*| `--jobname=`*jobname*	Job name. Required.
`-m` *npert2*| `--npert2=`*npert2*	Total number of perturbations to generate the complete second (free) FEP. Required.

Table 11.6. Options for the mcpro_ddg command. (Continued)

Option	Description
`-n` *npert1*| `--npert=`*npert1*	Total number of perturbations to generate the complete first (bound) FEP. Required.
`-t`|`--notrap`	Return complete error messages
`-v`|`--version`	Prints version of the `mcpro_ddg` program
`-w`|`--write`	Only write the specified input files

11.6 Linear Response Simulations

Linear response (LRM) simulations can be run with `mcpro_lrm`. This program can generate the input files including the required ligand and complex model systems from a prepared protein model system and a multi-structure ligand file in Maestro or SD format. You can also run a linear response simulation for a single ligand, provided that the bound and free input files and the model systems for the complex and ligand are available.

Syntax:

```
$SCHRODINGER/mcpro_lrm [options] -b infile -f infile2 -j jobname
   [-m ligand-file -p protein-file]
```

The options are described in Table 11.7.

Table 11.7. Options for the mcpro_lrm command.

Option	Description
`-b` *infile1*| `--bound=`*infile1*	MCPRO+ input file for the first (bound) perturbation. Required.
`-c`|`--compile`	Compile results from a completed LRM run only
`-f` *infile2*| `--free=`*infile2*	MCPRO+ input file for the second (free) perturbation. Required.
`-h`|`--help`	Display usage message and exit
`-i`|`--mini`	Minimize initial structures prior to running LRM
`-j` *jobname*| `--jobname=`*jobname*	Job name. Required. If you generated the input files from Maestro, you must use the job name that you used to write the files.
`-m` *ligand-file*| `--multilig=`*ligand-file*	Structure file of ligands to be processed, in Maestro or SD format.

Table 11.7. Options for the mcpro_lrm command. (Continued)

Option	Description
`-p` *protein-file*| `--prot=`*protein-file*	Maestro file with associated Z-matrix for protein (i.e. model system for protein).
`-q` *oplsvers*| `--oplsVersion=`*oplsvers*	Version of OPLS force field to be used in the simulations. Allowed values are 2001, 2005.
`-r`|`--rigSolute`	Constrain ligand degrees of freedom.
`-t`|`--notrap`	Return complete error messages
`-v`|`--version`	Prints version of the `mcpro_lrm` program
`-w`|`--write`	Only write the specified input files
`-y`|`--cmae`	Take ligand charges from the input ligand structure file.

11.7 Setting Up a Model System

The Python utility `mcpro_zmat` can be used to set up or modify a model system. This utility is used by Maestro to set up a model system.

A model system (protein, ligand, or complex) is required as input for all MCPRO+ jobs. For LRM jobs, only a protein model system is required, along with a multi-structure ligand file: the ligand-only and complex model systems are generated automatically as part of the job.

Syntax:

```
$SCHRODINGER/utilities/mcpro_zmat [options] -j jobname -s startligfile
   -e endligfile -p proteinfile
```

The options are described in Table 11.8.

Table 11.8. Options for the mcpro_zmat utility.

Option	Description
`-v`|`-version`	Show program version and exit
`-h`|`-help`	Display usage message and exit.
`-j` *jobname*| `--jobname=`*jobname*	Jobname for process that specifies output filenames
`-s` *startligfile*| `--ligstart=`*startligfile*	Starting ligand structure file in Maestro, SD, or PDB format.

Table 11.8. Options for the mcpro_zmat utility.

Option	Description
`-t` *cofactorfile*| `--cofactor=`*cofactorfile*	Non-peptide cofactors to be included in the Z-matrix.
`-e` *endligfile*| `--ligend=`*endligfile*	Ending ligand structure file in Maestro, SD, or PDB format.
`-p` *proteinfile*| `--prot=`*proteinfile*	Protein structure file in Maestro or PDB format
`-z` *protzmatfile*| `--protZmat=`*protzmatfile*	Protein Z-matrix in MCPRO format
`-r` *consfile*| `--rest=`*consfile*	Structure file used to define protein constraints/restraints in Maestro/SD format
`-c`|`--cap`	Add a solvent cap centered at either the constraint ligand or start ligand geometries.
`-x` *consspec*| `--consSpec=`*consspec*	Specify constraints to be used. 0=All flexible, 1=All rigid, 2=2 Shell, 3=3 Shell 4=2 Shell/Backbone fixed.
`-b` *dist*| `--consStart=`*dist*	Distance from the ligand specified by -r that full receptor constraints will be applied. Default: 8 Å.
`-o` *charge*| `--protCharge=`*charge*	Final desired protein charge. This value overrides the attempt by the script to determine the charge for neutrality.
`-q` *oplsvers*| `--oplsVersion=`*oplsvers*	Version of OPLS force field to be used in the simulations. Allowed values are 2001, 2005.
`-f` *dist*| `--cutsize=`*dist*	Residues within this distance from the ligand specified by `-r` will be kept for the model system. Default: 11 Å.
`-l` *logfile*| `--logFile=`*logfile*	Log file name. Default: *jobname*`.log`.
`-y`|`--cmae`	Take ligand partial charges from the input files instead of the force field.
`-d`|`--debug`	Run with debug printing on
`-a`|`--save`	Save temporary directory where model system is generated
`-n`|`--neut`	Run protein neutralization only
`-g`|`--chop`	Run protein chopping and capping only
`-w`|`--rigSolute`	Constrain ligand degrees of freedom
`-u`|`--flexCofactor`	Make cofactor fully flexible
`-i`|`--noChopNeut`	Do not perform chop/cap and neutralization for protein system

Table 11.8. Options for the mcpro_zmat utility.

Option	Description
`-bndTol=`*bndtol*	Tolerance for bonds to be included in Geometry Variation Section of Z-matrix. Default: 0.0001 Å.
`-angTol=`*angtol*	Tolerance for angles to be included in Geometry Variation Section of Z-matrix. Default: 0.001 degrees.
`-torTol=`*tortol*	Tolerance for torsions to be included in Geometry Variation Section of Z-matrix. Default: 0.001 degrees.

11.7.1 Output Files

Model system creation generates a Z-matrix file and a Maestro file for each part of the system: the protein, the ligands, and the complex. The files are listed in Section 3.7 on page 19. While the script is running, intermediate files are written to the subdirectory *basename*`_workdir`, This directory is deleted upon successful job completion. You can request that all intermediate files be saved by specifying `-save` and `-debug` (both required) on the command line.

11.7.2 Examples

Several examples of model system creation with notes are given below.

Ligand only model system creation with opls2005 force field and quantum charges:

```
$SCHRODINGER/utilities/mcpro_zmat -j ligOnly -s C02_start.mae
    -e C01_end.mae -c -q 2001 -y
```

`-j ligOnly`	Sets the job name and governs the names of the output files, which in this case are `ligOnly.log`, `ligOnly-lig.zmat`, and `ligOnly-lig.mae`. The two `ligOnly-lig.*` files constitute the model system.
`-s C02_start.mae`	Initial structure file for the MPCRO+ simulation. This structure file may be in Maestro or SD format
`-e C01_end.mae`	Final structure file for an MCPRO+ FEP simulation. If not performing an FEP set `-s` and `-e` options to the same file.
`-c`	Specifies that a solvent cap is to be placed at the center of the ligand.
`-q 2001`	Uses the OPLS2001 force field to atomtype the ligand structures.
`-y`	Indicates that atomic partial charges found in the Maestro file are to be used in place of the force field charges for the simulation. Charges generated by Jaguar are automatically added to the output Maestro file, which may be used as input to MCPRO+

Fully flexible protein-only model system creation with a chop radius of 12 Å:

```
$SCHRODINGER/utilities/mcpro_zmat -j protOnly
    -p 1cx2_prot_noheme.mae  -r C02_start.mae -c -x 0 -f 12
```

`-j protOnly`	sets the job name and governs the names of the output files which in this case are protOnly.log protOnly-prot.zmat and protOnly-prot.mae. The two protOnly-prot.* files constitute the model system.
`-p 1cx2_prot_noheme.mae`	sets the prepared protein structure to be used for a MCPRO+ simulation.
`-r C02_start.mae`	sets the reference structure about which protein chopping, neutralization, and flexibility options will be set.
`-c`	Specifies that a solvent cap is to be placed at the center of the ligand.
`-x 0`	All degrees of freedom of the protein will be treated flexibly. This is suited for a minimization run. If set to 1 all degrees of freedom would be fixed and if set to 2 a shell of residues about the reference structure (with a radius set by the `-b` option) will be treated flexibly with remaining residues fixed. A value of 2 is typically used for protein-ligand complex simulations. Backbone atoms are not flexible.
`-f 12`	Sets the distance from the reference ligand at which the input protein system will be chopped to 12 Å.

Protein-ligand complex creation typical for an FEP:

```
$SCHRODINGER/utilities/mcpro_zmat -j complex -s C02_start.mae
    -e C01_end.mae -p 1cx2_prot_noheme.mae -c -x 2 -q 2005 -f 12
    -y -r C02_start.mae
```

`-j complex`	Sets the job name and governs the names of the output files which in this case are `complex.log`, `complex-prot.zmat`, and `complex-prot.mae`. The two `complex-prot.*` files constitute the model system.
`-s C02_start.mae`	Initial structure file for the simulation. This structure file may be in Maestro or SD format.
`-e C01_end.mae`	Final structure file for the FEP simulation. If not performing an FEP set `-s` and `-e` options to the same file.

`-p 1cx2_prot_noheme.mae`	Sets the prepared protein structure to be used for the simulation.
`-c`	Specifies that a solvent cap is to be placed at the center of the ligand.
`-x 2`	Indicates that a shell of residues about the reference structure will be treated flexibly with remaining residues fixed. Side chains within 8 Å of the reference ligand will be flexible. Backbone atoms are not flexible.
`-q 2005`	Indicates that the OPLS2005 force field is to be used in atom-typing the ligand structure and any cofactors or metals found in the protein structure.
`-f 12`	Sets the distance from the reference ligand at which the input protein system will be chopped to 12 Å.
`-y`	Atomic partial charges found in the Maestro file are to be used in place of the force field charges for the simulation. Charges generated by Jaguar are automatically added to the output Maestro file which may be used as input to MCPRO+.
`-r C02_start.mae`	Specifies the reference structure about which protein chopping, neutralization, and flexibility options will be set.

11.8 Modifying the Input File When Reusing a Z-Matrix

If you want to reuse an existing Z-matrix, you must modify the `.inp` file before you run the simulation.

For a protein-ligand Z-matrix you will need to change, add, or remove the following lines for each parameter block in the input file (for the bound FEP only if you are running `mcpro_ddg`). Directions and comments on the status of the file written by Maestro are given next to the input file lines:

`ncent2=`*atomnum*	Change. *atomnum* is the atom number of first non-dummy atom in the ligand. This number is set to 0.
`nrot2=`*atomnum*	Change. *atomnum* is the atom number of first non-dummy atom in the ligand. This number is set to 0.
`isolec=2`	Add or change. This line is absent or the value is set to 1.

`adels2 = adels1` `rdels2 = rdels1`	Add for all calculations except minimizations. These are sampling options for the ligand. These lines are absent.
`noback=1`	Change for all calculations except minimizations, for which the value of 0 is correct. The value is set to 0 in the Maestro-generated file.
`nschg2=999999`	Remove any lines with this setting, except for minimizations.

For a ligand-only zmatrix you will need to remove any `nschg2=999999` lines from the input file, including the free FEP if you are running `mcpro_ddg`.

For a protein-only zmatrix you will need to remove any `nschg2=999999` lines from the input file, including the free FEP if you are running `mcpro_ddg`, and you will need to set `noback=1`.

11.9 Distributed Computing

MCPRO+ jobs can be distributed across multiple processors. For FEP simulations, the energy windows are distributed over the available processors. For relative binding affinities and linear response calculations, the processors are divided evenly between the bound and the free simulations, and then the energy windows or ligands are distributed over the available processors.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Maestro menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the tab that is displayed in a panel, click the Help button in the panel, or press F1. The help topic is displayed in your browser.
- For other information in the online help, open the default help topic by choosing Online Help from the Help menu on the main menu bar or by pressing CTRL+H. This topic is displayed in your browser. You can navigate to topics in the navigation bar.

The Help menu also provides access to the manuals (including a full text search), the FAQ pages, the New Features pages, and several other topics.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Maestro User Manual*, for detailed information on using Maestro
- *Maestro Command Reference Manual*, for information on Maestro commands
- *Maestro Overview*, for an overview of the main features of Maestro
- *Maestro Tutorial*, for a tutorial introduction to basic Maestro features
- MCPRO+ Frequently Asked Questions pages, at https://www.schrodinger.com/MCPRO+_FAQ.html
- Known Issues pages, available on the Support Center.

The manuals are also available in PDF format from the Schrödinger Support Center. Local copies of the FAQs and Known Issues pages can be viewed by opening the file `Suite_2009_Index.html`, which is in the `docs` directory of the software installation, and following the links to the relevant index pages.

Information on available scripts can be found on the Script Center. Information on available software updates can be obtained by choosing Check for Updates from the Maestro menu.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information:

- All relevant user input and machine output
- MCPRO$^+$ purchaser (company, research institution, or individual)
- Primary MCPRO$^+$ user
- Computer platform type
- Operating system with version number
- MCPRO$^+$ version number
- mmshare version number

On UNIX you can obtain the machine and system information listed above by entering the following command at a shell prompt:

```
$SCHRODINGER/utilities/postmortem
```

This command generates a file named *username*`-`*host*`-schrodinger.tar.gz`, which you should send to help@schrodinger.com. If you have a job that failed, enter the following command:

```
$SCHRODINGER/utilities/postmortem jobid
```

where *jobid* is the job ID of the failed job, which you can find in the Monitor panel. This command archives job information as well as the machine and system information, and includes input and output files (but not structure files). If you have sensitive data in the job launch directory, you should move those files to another location first. The archive is named *jobid*`-archive.tar.gz`, and should be sent to help@schrodinger.comIf Maestro fails, an

error report that contains the relevant information is written to the current working directory. The report is named `maestro_error.txt`, and should be sent to help@schrodinger.com. A message giving the location of this file is written to the terminal window.

More information on the `postmortem` command can be found in Appendix A of the *Job Control Guide*.

120 West 45th Street, 29th Floor
New York, NY 10036

101 SW Main Street, Suite 1300
Portland, OR 97204

8910 University Center Lane, Suite 270
San Diego, CA 92122

Zeppelinstraße 13
81669 München, Germany

Dynamostraße 13
68165 Mannheim, Germany

Quatro House, Frimley Road
Camberley GU16 7ER, United Kingdom

SCHRÖDINGER®